Fall 2014

Power-efficient high-speed interface circuit techniques

Elkim Felipe Roa Fuentes
Purdue University

Graduate School Form 30
(Revised 08/14)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Elkim Felipe Roa Fuentes

Entitled
Power-Efficient High-Speed Interface Circuit Techniques

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

BYUNGHOO JUNG

ANAND RAGHUNATHAN

KAUSHIK ROY

WILLIAM J. CHAPPELL

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

BYUNGHOO JUNG

Approved by Major Professor(s): _____

_____

Approved by: Venkataramanan Balakrishnan 09/25/2014

Head of the Department Graduate Program Date

POWER-EFFICIENT HIGH-SPEED INTERFACE CIRCUIT TECHNIQUES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Elkim Felipe Roa Fuentes

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2014

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

After long five winters -I mean long long-, it is time to look back and express gratitude to a great number of people who encouraged me and supported me during the journey in an exceptional place.

I am indebted to Professor Byunghoo Jung for granting me the privilege of being part of a team with outstanding technical qualities. Being able to interact with researchers coming from IIT, KAIST, SNU, NTU and NUS among other worldwide top-schools enriched my way of approaching technical problems with complete different perspectives. Professor Jung' entrepreneurialism and enthusiasm allowed me to explore and open doors in foreign subjects to the group's experience. And foremost, my gratitude for the philosophical sessions that we have held during few spare times. For passing one of my favorite and thought-provoking tales: the fisherman's parable.

Thanks to my labmates, specially Dr. Wu-Hsin Chen for the countless technical discussions. My gratitude to Srikar Bhagavatula for the life-related discussions after tireless tapeouts. Wing-Fai and Julia Lu for their assistance in layout when time was running faster than usual. Minsuk and Mohit for his soldering expertise. Bong Chan for his valuable help in the sign-off of two tapeouts. I also wish to acknowledge the continuous assistance with CAD tools of Dr. Mark Johnson. The several granted permissions of Dr. Matt Swabey that allowed me to move expensive instrumentation from ECE to Birck nanotechnology center.

My sincere appreciation to Kirk Arndt and Professora Daniela Bortoletto for letting me be part of the team that builds sensors for the large hadron collider. As soon as I knew that my first daughter was in her way, I started looking around on-campus for an additional income. At physics, I found a research assistant position that involved me for one year in the wonderful world of particles collision. Kirk besides of

teaching me the details of the sensors that brought the discovery of the Higgs, also taught me my first practical lessons of sailing.

I am bound to Mohan Chirala and Marko Aleksic who gave me the opportunity to interact with wonderful teams at Samsung and Rambus. At Samsung Dallas, I was fortunate to work with Kim Beumer who mentored my work during two consecutive tapeouts. At Rambus, special thanks to Pravin Kumar, Wendem Beyene, Shaobo Liu and Nikhil Vaidya for their technical discussions in the SERDES field.

My gratitude also goes to the professors who served in my committee. Professor William Chappell that took the time to assist to my defense even with his tight schedule as director of the microsystems technology office at DARPA. Professor Kaushik Roy and professor Anand Raghunatan for their feedback and suggestions.

Finally, it is time to mention the most important people. My wife, who get my deepest gratitude for her sacrifice and incredible patient love during the journey. I am more than life-thankful to my daughters Luciana and Keana. Although their birth during my graduate student time brought immense number of new challenges, their presence brought new life-perspectives and a boosted manner to live the day-by-day. Speaking of challenges, my sister Linda gets special thanks for spending weeks with my family while I was in ascetic mode during my last tapeout. My mama for inspiring me through her hard work to overpass the most difficult life challenges.

I appreciate the financial assistant of the sponsors of my research. In particular, the National Science Foundation, Samsung and the Electronics and Telecommunications Research Institute of South Korea. I am grateful to the Fulbright program sponsored by the U.S. Department of State for awarding me with a fellowship during my first years of my Ph.D. program.

This work is dedicated to my daughters Luciana and Keana because during playtimes I was able to articulate ideas with more fluency. To Luciana for reminding me every time that I was sitting for long-time in the desk: “Dadda no more busy, it is time to play”. To Keana for pulling me the legs under my desk when she was tired of crawling asking for some lap time.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Roa Elkim, Elkim Felipe Ph.D., Purdue University, December 2014. Power-Efficient High-Speed Interface Circuit Techniques. Major Professor: Byunghoo Jung.

Inter- and intra-chip connections have become the new challenge to enable the scaling of computing systems, ranging from mobile devices to high-end servers. Demand for aggregate I/O bandwidth has been driven by applications including high-speed ethernet, backplane micro-servers, memory, graphics, chip-to-chip and network on-chip. I/O circuitry is becoming the major power consumer in SoC processors and memories as the increasing bandwidth demands larger per-pin data rate or larger I/O pin count per component. The aggregate I/O bandwidth has approximately doubled every three to four years across a diverse range of standards in different applications. However, in order to keep pace with these standards enabled in part by process-technology scaling, we will require more than just device scaling in the near future. New energy-efficient circuit techniques must be proposed to enable the next generations of handheld and high-performance computers, given the thermal and system-power limits they start facing.

In this work, we are proposing circuit architectures that improve energy efficiency without decreasing speed performance for the most power hungry circuits in high speed interfaces. By the introduction of a new kind of logic operators in CMOS, called implication operators, we implemented a new family of high-speed frequency dividers/prescalers with reduced footprint and power consumption. New techniques and circuits for clock distribution, for pre-emphasis and for driver at the transmitter side of the I/O circuitry have been proposed and implemented. At the receiver side, new DFE architecture and CDR have been proposed and have been proven experimentally.

1. INTRODUCTION

During the last five years, the power density wall has transitioned from the microprocessor side to the peripheral side. Inter- and intra-chip connections have become the new challenge to enable the scaling of computing systems, ranging from mobile devices to high-end servers. Demand for aggregate I/O bandwidth has been driven by applications including high-speed ethernet, backplane micro-servers, memory, graphics, chip-to-chip and network on-chip. I/O circuitry is becoming the major power consumer in SoC processors and memories as the increasing bandwidth demands larger per-pin data rate or larger I/O pin count per component. Figure 1.1, adapted from the ITRS roadmap [4], shows that the aggregate I/O bandwidth has approximately doubled every three to four years across a diverse range of standards in different applications. However, in order to keep pace with these standards enabled in part by process-technology scaling, we will require more than just device scaling in the near future. New energy-efficient circuit techniques must be proposed to enable the next generations of handheld and high-performance computers, given the thermal and system-power limits they start facing.

1.1 High-speed links

The link performance depends on the required interconnection distance. Medium-distance electrical I/O links, such as server rack-to-rack and backplanes, must support high data rates along with lossy channels. Memory and chip-to-chip links must deal with crosstalk issues and channel loss in low-budget lossy channels. In most applications, there are common circuits that limit the power-budget of the electrical I/O links. From clocking circuits to driver circuits, including muxing circuits and equal-

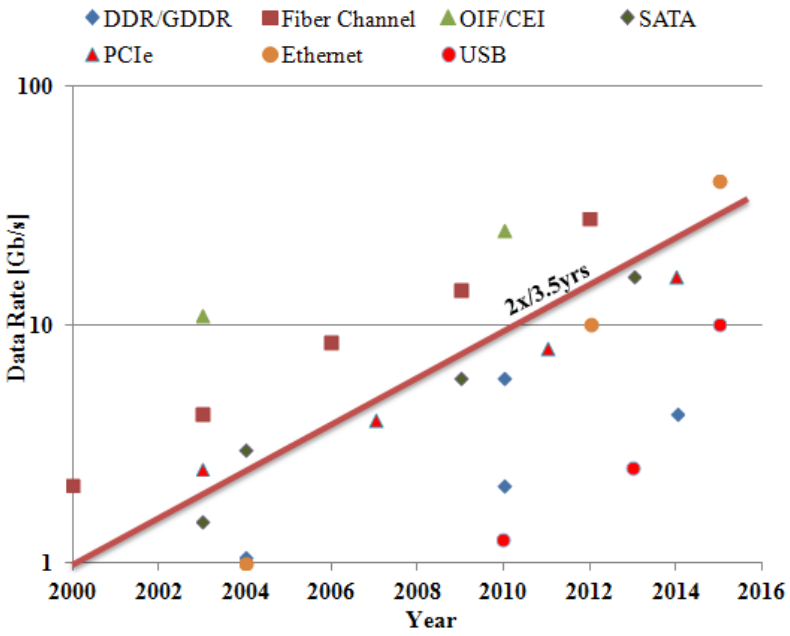


Fig. 1.1: Trend in per lane data rate for different wireline I/O standards.

izers at both ends (receiver and transmitter), efficient circuit topologies are required to keep the average power at the limits.

The block diagram of a traditional high-speed link is shown in figure 1.2. High-speed transceiver serialize parallel data to transmit it through a wireline or wireless channel. At the transmitter side, the clocking circuit scheme, a PLL and the frequency dividers are the timing generators in the link. The frequency dividers/prescalers are used inside the PLL at the feedback loop to serve as frequency multipliers. The frequency dividers also supply timing reference to the serializer blocks to allocate and space-synchronize parallel blocks of data onto a high-speed data stream.

In a wireline link, the most common channel is composed of PCB traces and connectors. Since the transient response of PCB traces at high frequency operation have frequency-dependent losses, the transceiver needs to include equalization circuits to compensate the channel-induced inter-symbol interference (ISI) at high data rates. These equalization circuits are commonly implemented at the output data rate to

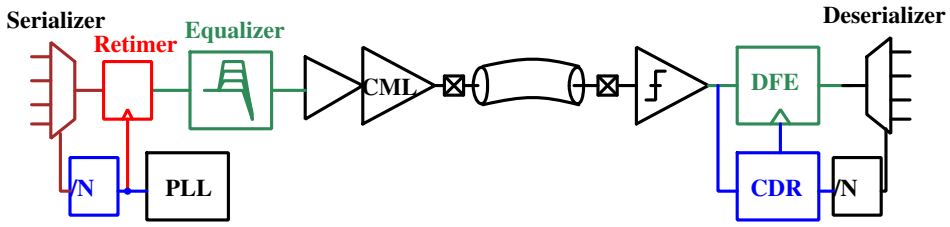


Fig. 1.2: Block diagram of a high-speed wireline transceiver.

have more control of the high-speed signal compensation. Therefore, the power consumption and overall performance of a high-speed link is strongly associated to the channel performance. Figure 1.3 reports the recent works presented at ISSCC and International VLSI Symposium. The energy efficiency values are presented over the last five years versus the channel loss at Nyquist frequency. The trend indicates about nine to ten times more power consumption in a lossy channel with respect to a channel with 20dB less loss at Nyquist frequency. The additional complexity required at the transmitter and receiver equalizers to compensate for several dB of channel loss explains this behavior.

Besides the equalization stage at the receiver side, other power-consumption dominant block is the clock-data recovery (CDR) circuit. The CDR recovers the clock and data from an amplifier usually followed by a continuous time linear equalizer (CTLE). The clock from the CDR is traditionally used in decision feedback equalizers that further compensate for ISI. The data from the CDR is sent to the deserializer that translates serial data on to parallel data.

1.2 Proposal

In this work, we are proposing circuit architectures that improve energy efficiency without decreasing speed performance for the most power hungry circuits in high-speed interfaces. For instance, the first frequency divider/prescaler seen by the high-speed clock is usually one of the limiting circuits in speed- and energy-budgets. By the

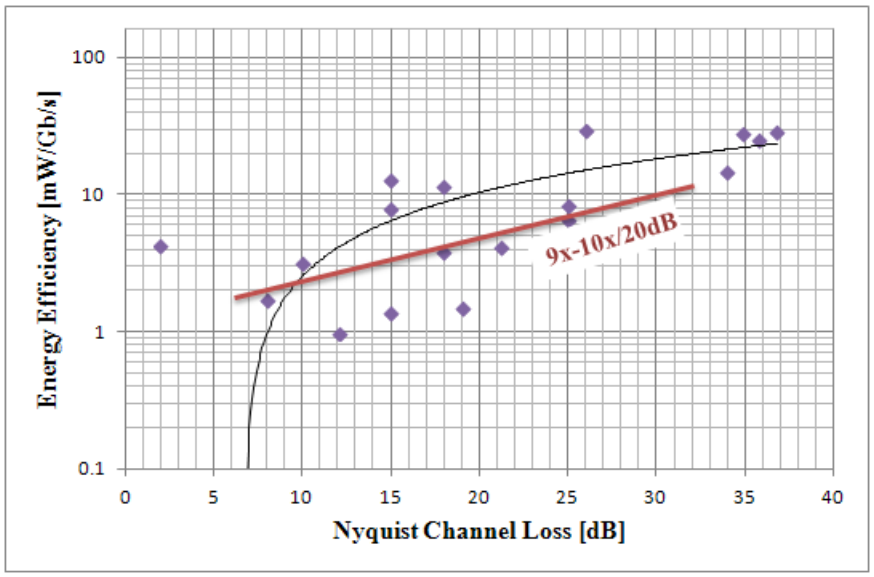


Fig. 1.3: Trend in wireline transceivers energy efficiency versus channel loss measured at Nyquist.

introduction of a new kind of logic operators in CMOS, called implication operators, we implemented a new family of high-speed frequency dividers/prescalers with reduced footprint and power consumption. In order to understand the framework used to design the new frequency dividers, chapter two introduces the implementation of the implication logic in CMOS. Chapter three focuses on the presentation of the family of new high-speed frequency dividers. Design and operation are described along with the experimental measurement results that demonstrated further performance improvement.

The transmitter circuitry is among the other limiting blocks in high-speed interfaces. The last stage of the muxing block, the pre-emphasis filter and the driver are commonly implemented with power-hungry circuit techniques to reach the desired operation frequency resulting a power budget constraint. Chapter four describes alternatives for the power hungry transmitter circuitry where experimental results indicate further energy-efficiency improvements compared to state-of-the-art transmitters.

Chapter five describes a proposed single-phase flip-flop capable to operate at 3.3GHz clock rate in a pipelined operation. Measurement results indicate a power savings over 40% when compared to the most common applied flip-flop in pipelined operations. New circuit techniques have also been proposed for the circuitry at the receiver side o wireline link. A new merged decision feedback equalizer and clock-data recovery topology has been implemented. Results show that the proposed scheme is energy-efficient when compared to reported work.

2. IMPLICATION LOGIC IN CMOS

We are used to computers with processors designed with millions of gates and memory cells, mostly, as a consequence of the semiconductor industry keeping up with Moore's law and the concern with shrinking familiar devices to increase the number (functionality) and speed by using the conventional switching logic framework. The switching logic basis was proposed by Shannon seventy years ago [5], and since then digital design has been using it. A quest for finding green devices has started considering that power density has become the barrier to continue further downscaling and enhancing of silicon devices. Therefore, most of the research has focused on finding new devices that can implement conventional switching logic.

In the 1950s, engineers started using transistors as devices to implement switching logic operations. However, transistors were invented initially for signal amplification in analog domain. As a result, the logic framework was proposed first, and later an existent device was reused from different domain to implement switching logic. Nowadays, new found devices have shown that new logic frameworks are required in order to take advantage of their natural operation.

A hundred years ago, the first volume of Whitehead and Russell's monumental work Principia Mathematica was published [6]. The book presented how the truths of math could be derived from logic operations. They described four fundamental logic operations, three of them used 27 years later by Shannon [5], and a fourth logic operation called material implication, A IMP B (also denoted as $A \rightarrow B$), which reads A implies B, and is equivalent to (NOT A) OR B. Russell and Whitehead emphasized the relevance of this logic operation by showing that only the IMP and NOT operations are enough to form a computationally complete logic basis. Because of the fact that Boolean logic operations, OR, AND, and NOT, form a computationally complete logic basis and they can be easily implemented using switching devices,

modern digital electronics have been founded on Boolean algebra, often referred to as switching algebra. In addition, the great achievement in modern electronics seems to indicate no need for extra logic functions. Consequently, the fourth fundamental logic operator, material implication (A IMP B), has been ignored during last a few decades and disappeared in many electronics textbooks (still plays an important role in logic theory textbooks). Three years ago, a letter in the journal Nature, presented that memristors can naturally execute the material implication operation, and inherently provides logic-in-memory [7]. However, these devices were fabricated using two layers of platinum-wire separated by an active layer of TiO_2, requiring a different process platform than current silicon process. Furthermore, reported speed performance was considerably low.

Meanwhile the search for the green device goes on, an alternative is to reconsider the logic framework using current devices instead of aiming changes in the device and the architecture direction. This work proposes a new kind of logic framework that can use existing charge-based devices more efficiently, or at least, extend current logic framework such that we can increase functionality around the power dissipation limits. This new logic framework uses different logic operations that can synthesize logic and memory operations with a reduced number of devices. Consequently, a given functionality would have less number of devices with greater energy efficiency.

In this research, we will show that CMOS circuits can also be used to perform material implications, IMP and NIMP (negated IMP). They are unit gates, which can extend or replace traditional Boolean logic operation. We present a different and promising paradigm based on material implication logic and CMOS circuits. Based on this new logic basis, we have demonstrated the faster, more compact and more efficient logic-based divider-by-2/3 prescaler. This comes as a result of the stateful logic nature of implication logic, and the opportunity that offers it to interpret digital logic in a distinct manner.

The truth table for the basic material implication operation is indicated in Fig. 2.1, meaning that if A is true, then the output after the implication operation will

Table 2.1. Truth table for IMP, $A \rightarrow B$.

A	B	B'
0	0	1
0	1	1
1	0	0
1	1	1

Table 2.2. Truth table for NIMP, A $\nrightarrow B$.

A	B	B'
0	0	0
0	1	0
1	0	1
1	1	0

be the value of B. The stateful nature of the implication operation, automatically provides a stateful logic operation if the operation can incorporate a memory. In general computing, logic and memory functions are required such that a logic variable stored in memory can be reused as input to another logic function. If a device or circuit is able to operate a logic gate and, a latch, it can provides a stateful logic capability. In [7], memristors were employed to provide a stateful logic using an unconventional framework using the material implication operation in a non-standard technology. The mining of this stafeul logic characteristic, allows to demonstrate that a new logic framework can be proposed.

2.1 Completeness of Implication Operators

In a two binary variables function, there are 2^2 possible arguments for each function. A set of functions is determined if we assign different values 1, or 0, to their 4 arguments. Therefore, 2^{2^2} functions are possible. A set of logic functions is called complete if and only if any possible logic function can be composed by a combination of functions of the set. A set of logical functions is complete if every logical function can be represented with members of the set. Post [8] points out the possibility of generalizing the work of Whitehead and Russell in Principia Mathematica by using other primitive functions instead of negation and disjunction. Post proved that a set of functions is complete if and only if the set contains at least one of each of the following types of functions:

1. non-linear: $f(x_1, x_2, ..., x_n)$ cannot be expressed in the form $(a_0 + a_1x_1 + a_2x_2 + ... + a_nx_n)$ where the a_n constants are either 0 or 1.

2. non-monotonic: There exist two states $(\hat{x_1}, \hat{x_2}, ..., \hat{x_n}), (\tilde{x_1}, \tilde{x_2}, ..., \tilde{x_n})$ such that $\hat{x_1} \leq \tilde{x_1}$, $\hat{x_2} \leq \tilde{x_2}$, ..., $\hat{x_n} \leq \tilde{x_n}$, and $f(\hat{x_1}, \hat{x_2}, ..., \hat{x_n}) > f(\tilde{x_1}, \tilde{x_2}, ..., \tilde{x_n})$.

3. non-self dual: $f(\hat{x_1}, \hat{x_2}, ..., \hat{x_n}) \neq f(\tilde{x_1}, \tilde{x_2}, ..., \tilde{x_n})$ for at least one state.

4. non-zero preserving: $f(x_1, x_2, ..., x_n) = 1$ for $x_1, x_2, ..., x_n = 0$.

5. non-unity preserving: $f(x_1, x_2, ..., x_n) = 0$ for $x_1, x_2, ..., x_n = 1$.

The functions NAND and NOR are functions that meet the previous properties by themselves. Consequently, the functions NAND and NOR are the only complete functions, called sometimes the universal logic primitives. That is, the NAND function can be used to compose every logical function, same thing we can do with the NOR function. Using the previous properties is possible to prove that the following set of pairs are complete: $\{\rightarrow, \perp\}$ (IMP,FALSE), $\{\rightarrow, \neg\}$ (IMP, NOT), $\{\nrightarrow, \neg\}$ (NIMP, NOT) , $\{\nrightarrow, \top\}$ (NIMP, TRUE), $\{\rightarrow, \nrightarrow\}$ (IMP, NIMP), $\{\nrightarrow, \leftrightarrow\}$ (NIMP, XNOR), $\{\rightarrow, \not\leftrightarrow\}$ (IMP, XOR), $\{\wedge, \neg\}$ (AND, NOT), and $\{\vee, \neg\}$ (OR, NOT). In fact, by showing that the NAND or the NOR operation are either members of each of the previous nine two-function sets, it is possible to say that the sets are complete. As a consequence, we can take any pair and develop a complete logic system with their respective algebra or calculus propositional. Interestingly, the operation cells, IMP and NIMP, are common in most of the pairs. For instance, we develop an axiom system for the sets $\{\rightarrow, \nrightarrow\}$, $\{\rightarrow, \neg\}$, $\{\rightarrow, \perp\}$ in the next section.

2.2 Logic operations with implication cells

Implementing the IMP and NIMP $\rightarrow, \nrightarrow$ operations in such a way that they behave as memory devices at the same time, enables them to recursively synthesize all the logic operations, as shown in Table 2.3. Similarly, it can be shown for the sets

Table 2.3. Logic operations with IMP and NIMP, $(\rightarrow, \nrightarrow)$.

NOT A	$A \rightarrow (A \nrightarrow A)$
A OR B	$(A \rightarrow (A \nrightarrow A)) \rightarrow B$
A NOR B	$(A \rightarrow (A \nrightarrow A)) \nrightarrow B$
A AND B	$A \nrightarrow (A \nrightarrow B)$
A NAND B	$A \rightarrow (A \nrightarrow B)$
A XOR B	$(A \rightarrow B) \rightarrow (B \nrightarrow A)$
A XNOR B	$(A \rightarrow B) \nrightarrow (B \nrightarrow A)$

$\rightarrow, 0$ and $\rightarrow, \neg$ as presented in Table 2.4 and Table 2.5 respectively. Assuming that exist different structures in CMOS that can perform the material implication and at the same time they behave as a memory devices, the question is how many of these structures[1] are required to compute the Boolean functions, and how they are interconnected.

Definition 1 *Let $M = \{m_1, m_2, ..., m_n\}$ be a set of memMOS. Each memMOS can perform the FALSE operation, set a 0, and the material implication operation. If for any Boolean function B exists a sequence of minimum number of memMOS such that m_n computes B, we can claim that the set of n memMOS is the smallest number of m functionally complete or universal.*

Theorem 1 *$M = \{m_1, m_2, m_3\}$ is universal.*

Proof One way to prove this is looking for a sequence of computations that can perform a NAND operation using just three memMOS. Let perform the operation $\neg(A \wedge B)$ (A NAND B) setting the value of A in m_1, the value of B in m_2, and the result s in m_3. Executing the following sequence results in the operation NAND:

$m_3 = 0, m_2 \rightarrow m_3, m_1 \rightarrow m_3$

the three steps sequence is indicated in table 2.6, where final result in m_3 (s'' is the result of A NAND B. ■

Table 2.4. Logic operations with IMP and FALSE, $(\rightarrow, 0)$.

NOT A	$A \rightarrow 0$
A OR B	$(A \rightarrow 0) \rightarrow B$
A NOR B	$((A \rightarrow 0) \rightarrow B) \rightarrow 0$
A AND B	$(A \rightarrow (B \rightarrow 0)) \rightarrow 0$
A NAND B	$A \rightarrow (B \rightarrow 0)$
A XOR B	$(A \rightarrow B) \rightarrow ((B \rightarrow A) \rightarrow 0)$
A XNOR B	$((A \rightarrow B) \rightarrow ((B \rightarrow A) \rightarrow 0)) \rightarrow 0$

Table 2.5. Logic operations with IMP and NOT, $(\rightarrow, \neg)$.

NOT A	$\neg A$
A OR B	$(\neg A) \rightarrow B$
A NOR B	$\neg((\neg A) \rightarrow B)$
A AND B	$\neg(A \rightarrow (\neg B))$
A NAND B	$A \rightarrow (\neg B)$
A XOR B	$(A \rightarrow B) \rightarrow (\neg(B \rightarrow A))$
A XNOR B	$\neg((A \rightarrow B) \rightarrow (\neg(B \rightarrow A)))$

A different way to prove is showing that exists sequences using $M = \{m_1, m_2, m_3\}$ such that the logic operations OR , NOT, and AND can be performed:

Proof Let set the value of A in m_1 and B in m_2 and be m_3 a memMOS working as temporal latch. Then the logic operations can be performed using the following sequences:

i. $m_3 = 0, m_1 \rightarrow m_3, m_3 \rightarrow m_2$. Operation $A \vee B$ with output stored in m_2 or $m_2 = (A \rightarrow 0) \rightarrow B$.

ii. $m_3 = 0, m_1 \rightarrow m_3$. Operation NOT $(\neg A)$ or $m_3 = A \rightarrow 0$.

iii. $m_3 = 0, m_2 \rightarrow m_3, m_1 \rightarrow m_3, m_2 = 0, m_3 \rightarrow m_2$. Operation $A \wedge B$ or $m_2 = (A \rightarrow (B \rightarrow 0)) \rightarrow 0$.

[1]From now we will called those structures: memMOS.

Table 2.6. Function NAND performed as a sequence with three memMOS.

$m_3 = 0$	$m_2 \rightarrow m_3$			$m_1 \rightarrow m_3$		
s	B	s	s'	A	s'	s''
0	0	0	1	0	1	1
0	0	0	1	1	1	1
0	1	0	0	0	0	1
0	1	0	0	1	0	0

■

Besides completeness of implication operators, another measure that indicates their promising applicability to improve logic functions is the number of gates needed to express all binary operators, called expressiveness. In [9], a method to measure the expressiveness was applied to a set of different minimal-universal libraries. Results showed that the set (IMP, NIMP) outperform over the functions NAND, NOR and the set of different universal pairs. Considering that the (IMP, NIMP) functions are more expressive, make them interesting operators for logic synthesis.

2.3 Basic axioms

Propositional logic was a hot subject in the first two decades of the 1900s. A large number of mathematicians and logicians were looking for a minimal axiom set that defines an universal formal system capable to derive the whole true propositions [10]. Lucasiewicz in 1917 formulates an axiomatic system with the following three axioms [11]:

i. $A \rightarrow (B \rightarrow A)$

ii. $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$

iii. $(\bar{A} \rightarrow \bar{B}) \rightarrow ((\bar{C} \rightarrow B) \rightarrow C)$

where the operation $\rightarrow$ can be read as logically implies. With these three axioms based on the set $\rightarrow, \neg$ is possible to derive a infinite number of axioms. A set of other useful axioms that can be applied to simplify circuits based on implication logic are the following:

i. $(\bar{A} \rightarrow \bar{B}) \rightarrow (B \rightarrow A)$

ii. $A \rightarrow A$

iii. $(\bar{A} \rightarrow A) \rightarrow A$

iv. $(A \rightarrow B) \rightarrow A \equiv A$

v. $B \rightarrow C, C \rightarrow D \vdash B \rightarrow D$

vi. $(A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C))$

vii. $(A \rightarrow (B \rightarrow C)) \equiv (B \rightarrow (A \rightarrow C))$

viii. $A \nrightarrow 0 \equiv B$

ix. $(A \nrightarrow B) \rightarrow A \equiv A$

x. $B \nrightarrow (A \nrightarrow B) \equiv B$

xi. $1 \rightarrow A \equiv A$

xii. $1 \nrightarrow A \equiv \bar{A}$

xiii. $A \rightarrow (B \rightarrow \bar{A}) \equiv B \rightarrow \bar{A}$

xiv. $(A \nrightarrow B) \nrightarrow B \equiv A \rightarrow B$

xv. $A \rightarrow \bar{B} \equiv B \rightarrow \bar{A}$

2.4 Implication cells in CMOS

The previous section presented a new complete approach using the material implication functions to synthesize Boolean operations. This section focus on the implementation of the implication functions in CMOS, IMP and NIMP. Initially, IMP and NIMP implemented without memory and working as a static CMOS gates are presented. Next, Implication operations that include memory are presented and their stateful functionality is described.

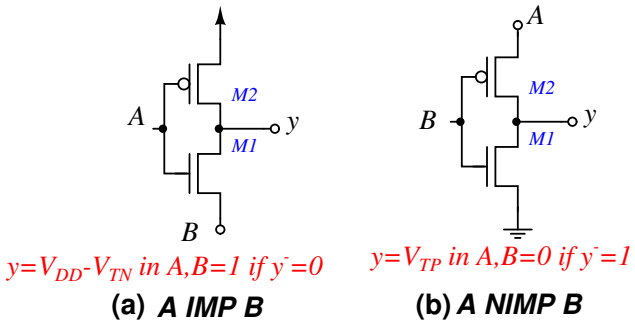


Fig. 2.1: (a) IMP cell (b) NIMP cell.

In order to see how the implication operation can be included in fully combinatorial logic, static cells are implemented in CMOS. A reduced IMP function is implemented in Fig. 2.1a. The IMP cells consists of a CMOS pair with one of the inputs connected to the source of the nMOS transistor. A limitation of this configuration is the weak high logic value at the output when both inputs are high and the previous value was the logic value zero. However, the configuration achieves comparable energy efficiency to a CMOS inverter considering that both transistors are operating in saturation in a short period of time when both inputs are transitioning values. Simulation results for different input patterns are shown in Fig. 2.2, where the output has been colored with different red tones. The dark-red tone indicates that the nMOS transistor is operating in saturation. The red and the light-red colors indicate that the nMOS transistor is operating in triode and sub-threshold respectively.

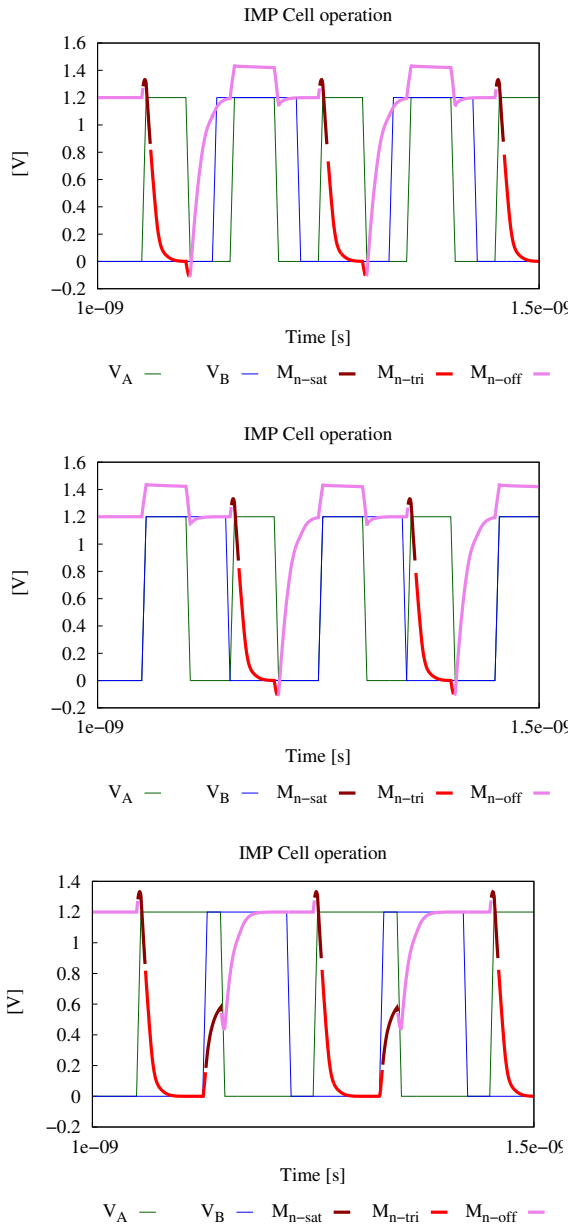


Fig. 2.2: Simulated waveforms of the IMP cell for different input patterns.

The static NIMP cell at transistor level is presented in Fig. 2.1b. Like the IMP cell, for a combination of the inputs, the output has a weak value. When the both inputs are zero logic, and the previous output was the high value logic, the low value output does not reach the ground level instead the value reaches $|V_{TP}|$. Similarly, the

static current is reduced since both transistors operate in saturation at shorter time periods. Simulation results for a 10 GHz period in one of the inputs is presented in Fig. 2.4.

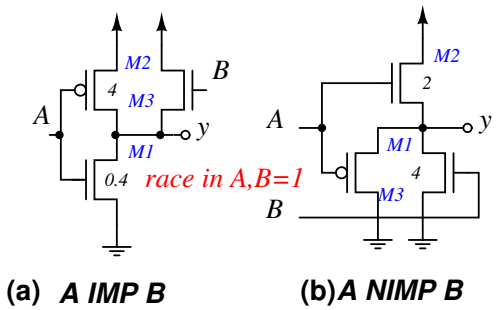


Fig. 2.3: (a) IMP cell (b) NIMP.

2.5 IMP and NIMP Cells with memory

The traditional digital wisdom is to use a pair of transistors as an inverter, or combine a pair with a resistor to obtain a switching logic circuit as it was proposed by Shannon. For instance, circuit designers have been using the equivalent NAND architecture as Shannon described it by using relays and Boolean logic. Fig. 2.5 shows different circuit techniques used to replicate the operation NAND using CMOS transistors. Figure 2.5(a) presents the initial NAND idea presented by Shannon using relays. Notice the different equivalent circuits used in current VLSI circuits using static and dynamic gates shown in 2.5. Each circuit family has your advantages and disadvantages but all of them are implemented using the original switching concept.

A latch that can execute the implication operation between its input and output can synthesize Boolean logic operations. As it was described in the logic framework, recursive logic gates can be synthesized with implication operators that have the ability to store a variable. For instance, Fig. 2.6 shows a flip-flop with the capacity to execute the IMP operation. An initial low level at the output and the value of

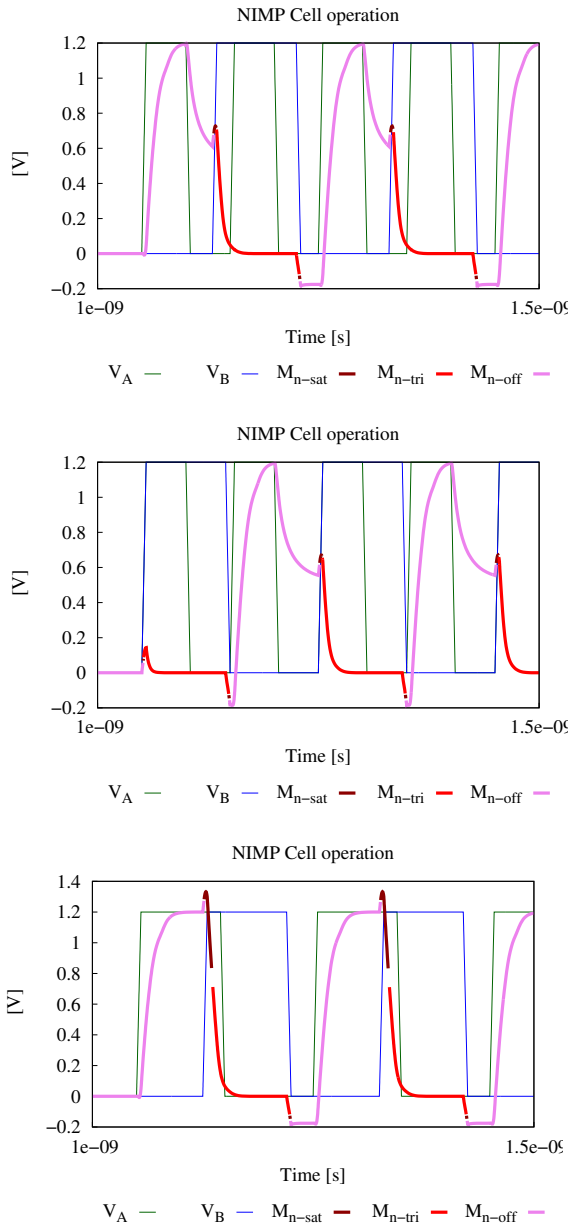


Fig. 2.4: Simulated waveforms of the NIMP cell for different input patterns.

a variable A at the input of the flip-flop, will output at the first clock-edge $A \rightarrow 0$, which is equivalent to $\overline{A}$. At the second clock-edge, the output of the flip will be the operation $B \rightarrow \overline{A}$, which is equivalent to A NAND B. As a result, the output of the flip-flop works as one of the inputs of the operation IMP, or to better understand

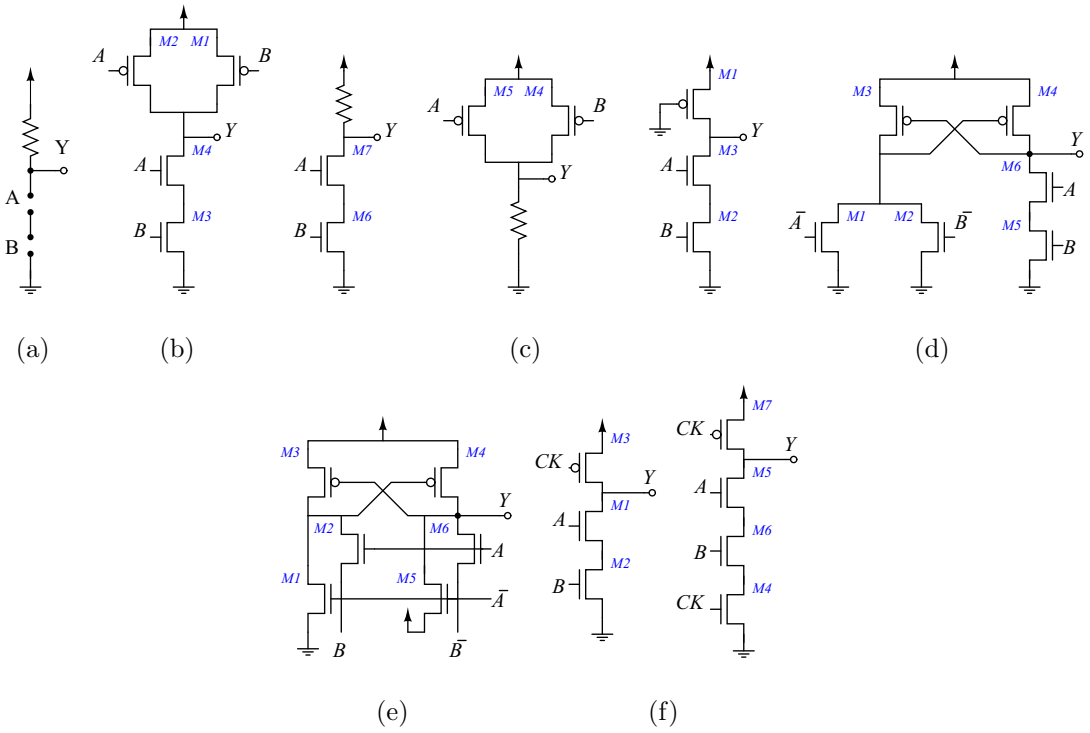


Fig. 2.5: NAND implementation in different CMOS families. (a) Switching logic concept for NAND from Shannon. (b) CMOS static. (c) Pseudo-nMOS and pseudo-pMOS. (d) Differential Cascode Voltage Switch Logic, DCVSL. (e) DCVS with pass transistors. (f) Dynamic logic.

it, it can be interpreted as one operand of IMP. The result of the operation IMP will be stored at the output of the flip-flop. This configuration allows to execute the operation NAND in two clock cycles and have their result in a memory device without having logic+memory penalties at the interconnection.

Figure 2.7 shows an AND pipelined operation using a NIMP + flip-flop cell. The within NIMP operation at the first clock-edge executes $A \nrightarrow 0$, which is equivalent to A. At the second clock-edge the operation $B \nrightarrow A$ is evaluated and stored. At the third clock-edge $B \nrightarrow (B \nrightarrow A)$ is executed and its result is equivalent to A AND B. This cell configuration synthesizes the AND operation in three clock cycles. Transistor level implementations of these cells are described in the application section.

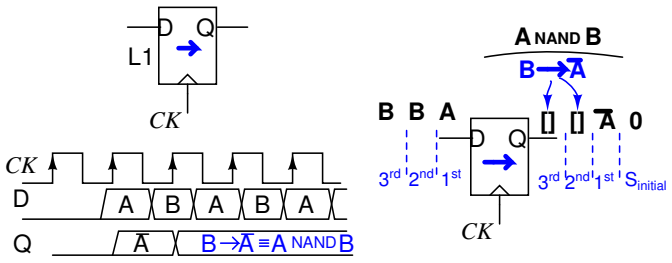


Fig. 2.6: (a) IMP + memory pipeline operation.

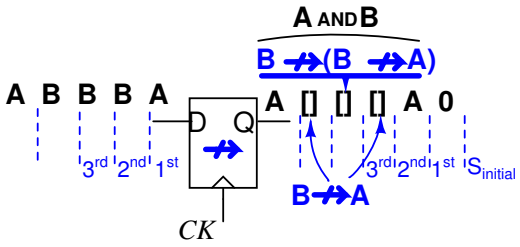


Fig. 2.7: (a) NIMP + memory pipeline operation.

2.5.1 Polymorphic NIMP and IMP Cells

We focus on the execution of material implication operation in simplest CMOS circuits. Considering the property of each memMOS to work as a memory device or latch, and the goal of finding the simplest circuit able to mimic the material implication operation, TSPC flip-flops [12] are used to implement the memory device and the implication operation. Fig. 2.8 shows a TSPC flip-flop where its output is *overlineQ*, the negated input after one cycle. This operation can be seen as a NOT device with memory. Recalling the logic framework, if we can implement the IMP or the NIMP operation with a similar circuit, we Will be able to perform all Boolean operations with those circuits. Moreover, if we will be able to implement both operations with just one circuit and have control to decide which operation to perform, pipelined logic functions can be operated within the same circuit.

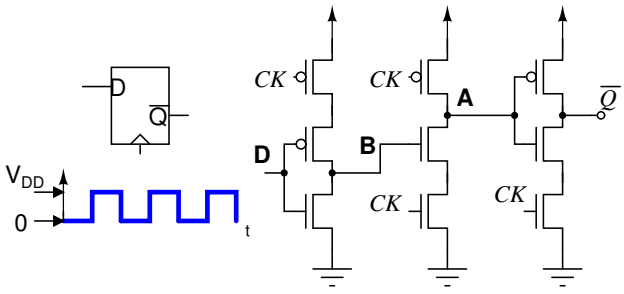


Fig. 2.8: TSPC flip-flop.

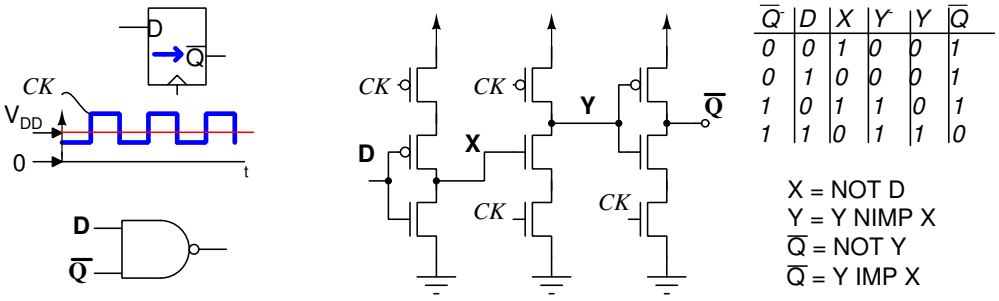


Fig. 2.9: TSPC flip-flop with polymorphic operation.

A flip-flop that can perform the IMP is described in Fig. 2.9. If the clock offset is shifted up, the circuit works as a memory and IMP cell. It operates the function $D \rightarrow Q$ storing the output at $\overline{Q}$ node. This operation is equivalent to D NAND $\overline{Q}$. If the clock voltage is not shifted, the flip-flop works as data inverter and memory cell. Fig. 2.9 presents a true table with the different values at the internal nodes where $\overline{Q^-}$ and Y^- are the previous values of $\overline{Q}$ and Y. Notice that between nodes X and Y, the function NIMP is performed. This behavior will be used to describe the frequency prescalers presented in next chapter.

3. FREQUENCY PRESCALERS USING IMPLICATION LOGIC

Clock generation synthesis consumes significant power in wireless and high-speed wireline applications. Power dissipation in PLL strongly depends on VCO and frequency dividers due to the number of devices switching at high frequencies. High-speed serializers are mostly limited in power and maximum data rate operation by the first frequency divider seen by the output mux. In a power saving strategy, the first frequency prescaler which sees the VCO output frequency is a key building block to be considered to decrease overall power.

A widely used Common Mode Logic (CML) divider topology is based on differential amplifiers like the one shown in figure 4.2. Using this approach fixed-division-ratio prescaler have achieved maximum operating frequencies up to 90GHz [1], at the expense of complementary clock phases generation and power dissipation. CML dividers require a higher number of stacked devices and a constant DC bias current which demands a considerable power consumption. Moreover, CML circuits require complementary clock phases making them vulnerable to skew and overlapping problems.

The constant scaling and thereby improvement of the intrinsic speed in CMOS devices, makes it possible to use true-single-clock-phase (TSPC) logic gates as a successor to CML in relative high-speed applications. However, TSPC-based dividers have been limited to applications below 20GHz due to the number of devices connected at high-speed nodes. In this paper, we have demonstrated a simplified divider-by-2/3 operating at 50GHz. We propose a new dual-modulus prescaler scheme using extended-TSPC [13] logic with a reduced number of transistors allowing high-frequency division. Compared to conventional TSPC and ETSPC prescaler schemes, the proposed prescaler achieves higher power efficiency, smaller footprint and higher maximum-

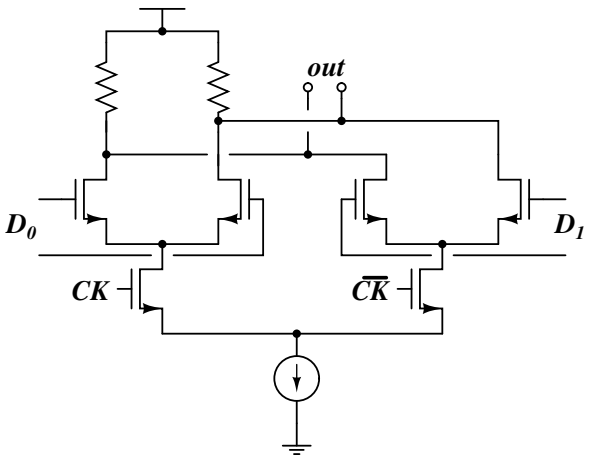


Fig. 3.1: CML static frequency divider.

frequency operation. The advantage of our approach becomes most apparent when compared to the work in [3], which we improved with a minimalist divider-by-2/3 prescaler. Following this introduction, section II provides a survey and analysis of frequency prescaler techniques. Section III presents the new prescaler along with section IV describing the obtained results.

3.1 High-speed frequency prescalers

Frequency plans in high-speed multi-standard transceivers requires highly-programmable frequency divider-ratio in their frequency synthesizers to perform adequate band tuning. Consequently, dividers at the first stage of the prescaler require to have a dual-modulus ratio. First stage divider is a challenging block considering the high-frequency operation and thereby the elevated power consumption required.

Traditional high-speed frequency dividers operating above the upper-half centimeter-wave band, are based on CML circuits and injection-locked topologies [1,14–16]. Conventional CML-based dividers use a CML DFF (D-Flip-Flop) with a feedback path, as the one presented in Fig. 4.2. In order to reduce rise and fall times, tuned in-

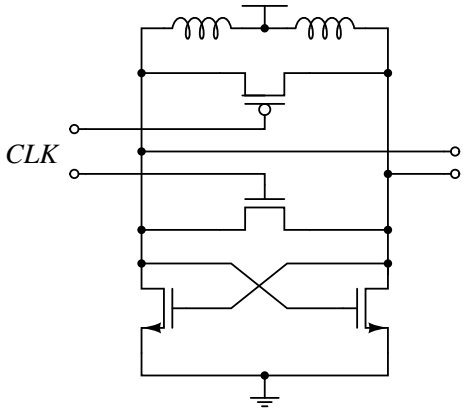


Fig. 3.2: Injection-locked frequency divider-by-2 in [1].

ductors are used to enhance the maximum input frequency. However, CML latches exhibit a larger node capacitances and high headroom demanding more power. Multiple techniques have been employed to reduce power consumption, such as, resonant tank loads [14], and capacitive-bridged shunt peaking technique [15], Fig. 4.2. Recently, more power efficient dividers using injection locking techniques have been reported [1, 16], as the one shown in Fig. 3.2. Nevertheless, division factors different than 2 are limited, and power consumption is in the range of few mWs which still prohibitive in the lower-half of the centimeter-wave band.

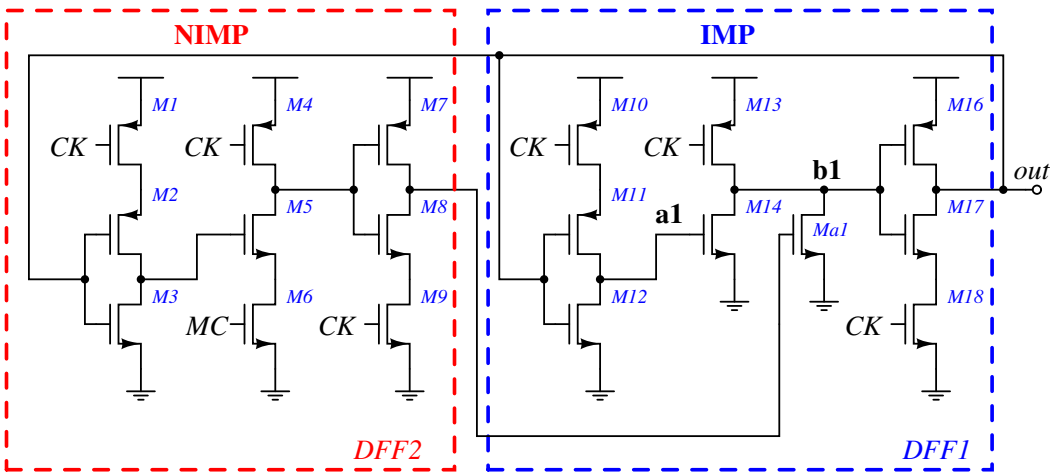
A common approach to implement frequency dividers in the first half of the centimeter-wave band, is to use TSPC logic structures [2,3], whereby a single clock is required to drive the logic. Thus, there is not need for a differential clock generation as in CML, and skew/overlapping issues do not take place. TSPC cells use three transistors in each branch, as shown in Fig. 3.3, while ETSPC cells use two transistors for branch, Fig. 3.4. Consequently, ETSPC cells reduce the number of transistors driven by the clock, requiring reduced-size transistors for the clock drivers and for some internal nodes of the DFFs. Intuitively, one can conclude that ETSPC cells operate at higher frequencies than TSPC cells due to the reduced load capacitance. However, ETSPC cells have a time period where a direct current-path exists between

rails. Therefore, one also intuitively says that ETSPC cells are more power hungry than TSPCs. On the other hand, switching power dissipation (P_w) is a key factor in the overall power, and is given by the sum of switching power at each node in equation 3.1,

$$P_{sw} = \sum_{i=1}^{n} f_i C_{L_i} V_{DD}^2 \tag{3.1}$$

where n is the number of switching nodes, f_i is the effective switching frequency of node i, C_{L_i} is the load capacitance at the output node of each ith-stage, and V_{DD} the supply voltage. Therefore, considering the reduced load capacitance in ETSPC cells, and the reduced number of stacked transistors that allows lower voltage supply, there is not consistent statement to conclude in a general manner that ETSPC-based prescalers consume more power than TSPC-based ones. It is not a fair to compare TSPC and ETSPC cells at the same frequency [2, 17] and/or same voltage [3], considering that at higher frequencies the switching power increases linearly, leaving the TSPC in disadvantage; and at low frequencies the static power increases over the switching power for the ETSPC case. Moreover, ETSPC cells can have reduced voltage supply considering the reduced number of transistors stacked. A more conclusive approach to compare the cells is to use the power-efficiency in mW/GHz taking in consideration minimum voltage supply. In summary, applications aiming the upper centimeter-wave upper band and the mm-wave band using high-f_t CMOS devices, should consider ETSPC as a promising technique to implement low power frequency dividers. Applications aiming frequencies below the half centimeter-wave band with high-f_t devices, should explore the use of TSPC cells for low power.

A key point to achieve high-frequency operation in TSPC and ETSPC cells is to reduce the capacitance loads in the critical path. Different techniques have been proposed to include the logic gates within the flip-flops in order to reduce the number of stages in cascade and therefore, reducing the path delay. However by doing this, capacitance loads increase at the nodes, therefore, requiring greater drive capability in internal nodes.



A	B	A NIMP B
0	0	0
0	1	0
1	0	1
1	1	0

A	B	A IMP B
0	0	1
0	1	1
1	0	0
1	1	1

Fig. 3.3: Divider-by-2/3 prescaler proposed in [2].

A novel divider-by-2/3 with ETSPC cells is presented in [3], which achieved the highest input frequency reported in TSPC-based prescalers. Authors in [3] present a clever idea which includes the OR and AND gates in the same branch, Fig. 3.4. However an inverter between the flip-flops adds a stage to the logic chain, limiting the maximum frequency. Additionally, in divider-by-2 mode (MC=0), the last two branches of DFF2 still switching, as indicated in figure 3.5 which burns power unnecessarily. This is the reason to the greater consumption reported in ETSPC-based compared to the ratioed-TSPC-based prescaler.

In [2] is presented a divider-by-2/3, Fig. 3.3, which uses TSPC cells. The maximum input frequency achieved is 8GHz in a 0.13μm process node. Although there is not additional inverter or logic gate between the flip-flops, the logic operation is realized in the critical path. Transistors Mc1 and Ma1 load the node b1, reducing the maximum frequency achievable. Besides using three transistors for branch, the

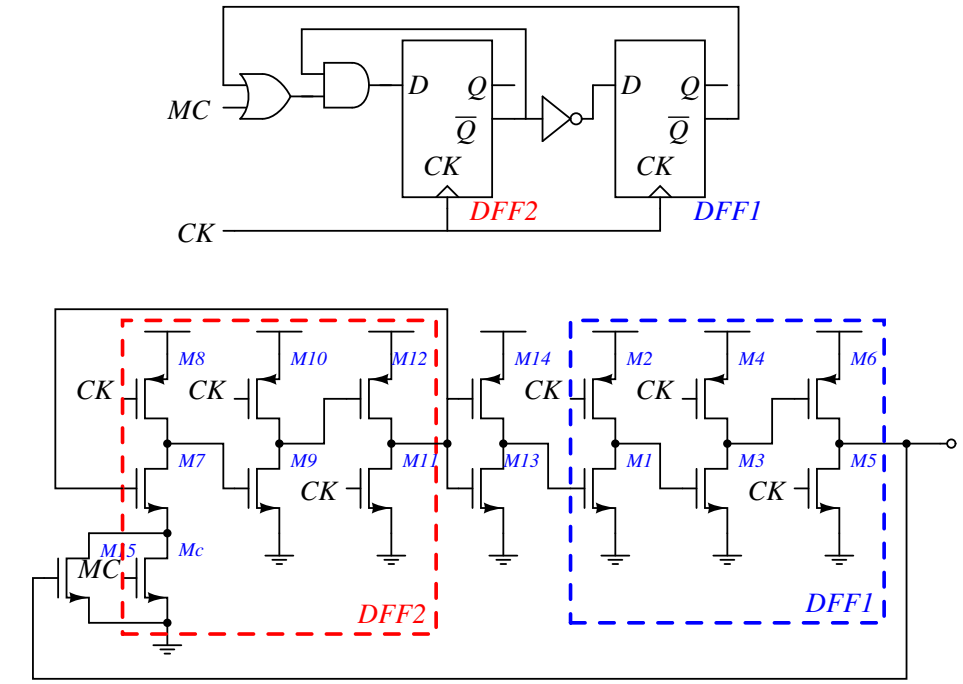


Fig. 3.4: Divider-by-2/3 prescaler proposed in [3].

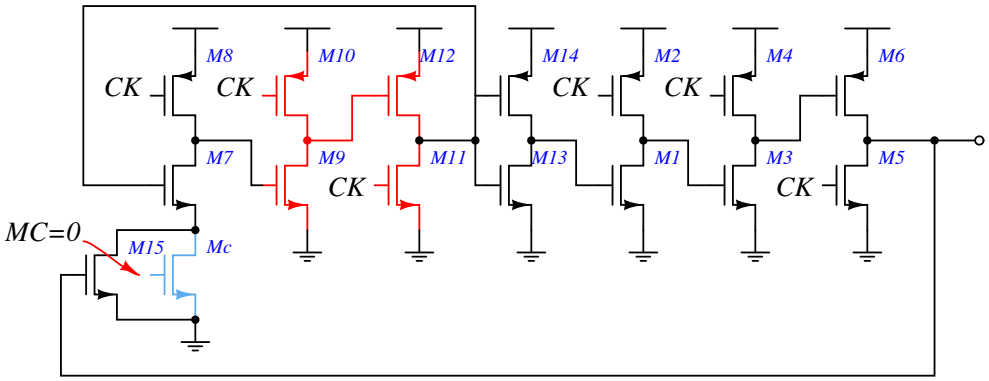


Fig. 3.5: Divider-by-2 operation in [3] prescaler.

added stacked transistors Mc1 and Ma1 to the node b1, force to increase the size of M13 which implies more load capacitance to node a1 and b1.

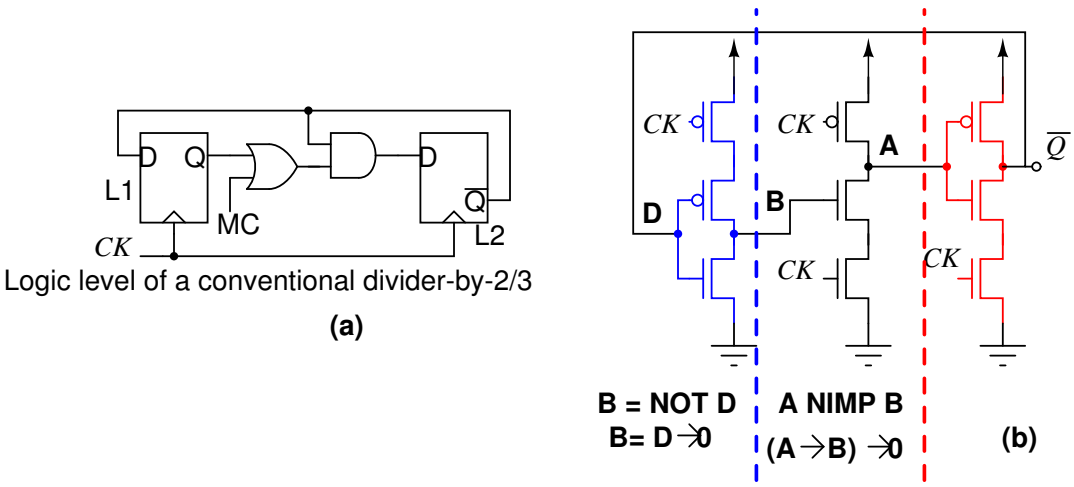


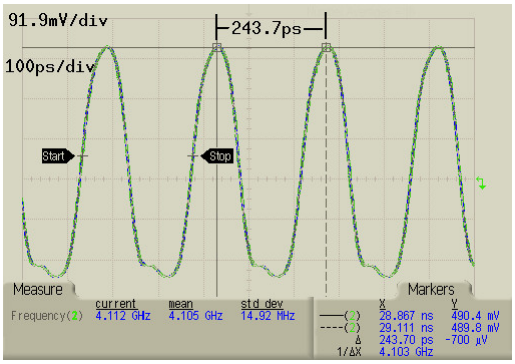
Fig. 3.6: (a) Conventional frequency divider-by-2/3. (b) Frequency divider-by-2/3 using NIMP operation.

3.2 Compact frequency prescalers using implication operators

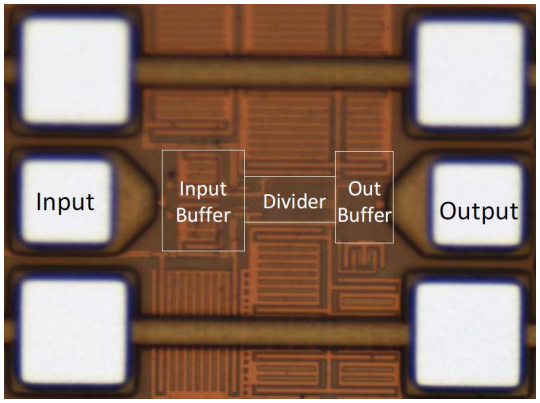
The operations NAND, NOR, OR, AND and NOT, have been used so far as the key gates to synthesize most of the digital logic in CMOS. However IMP and NIMP operators, can provide a shortcut function to optimize certain logic operations at circuit level. We present a dual-modulus divide-by-2/3 shown in gate level at Figure 3.8a as an example. This is an appropriate example considering it involves combinatorial and sequential operators. This circuit is commonly used in frequency synthesizers where speed performance and power consumption are key features. The conventional wisdom is to synthesize the circuit using the traditional switching logic concepts proposed by Shannon. For instance, recently published prescalers have proposed to add the AND and OR logic gates within the flip-flops to reduce the number of stages in cascade, reducing the path delay [17]. However, in this approach, parasitic capacitances at internal nodes increase, requiring greater current driving capability and consequently more power consumption and slower speed.

Table 3.1. NIMP operation explained.

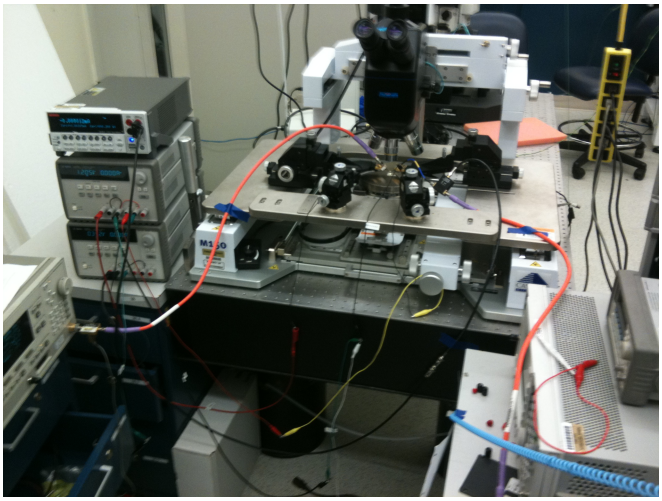
A	B	$A_{CK=0}$	$A_{CK=1}$	$A_{CK=1+}$	$A_{CK=0+}$
0	0	1	0	0	0
0	1	1	0	0	0
1	0	1	1	1	1
1	1	1	0	0	0



(a)



(b)



(c)

Fig. 3.7: (a) Measured output in divide-by-3 operation mode. (b) Microphotograph of divider-by-2/3.(c) Measurement setup for divide-by-2/3 prescaler.

3.3 Using polymorphism on Flip-Flops

The logic gate that includes AND and OR can by replaced by a dynamic NIMP to provide the same functionality. The NIMP operation is achieved by shifting the DC level of the clock signal connected to the flip-flop shown in Figure 3.8b. By doing this, the operation performed at node A is $(A \rightarrow B) \rightarrow 0$, which is the same as A NIMP B. The complete performed operation within the flip-flop is explained in Table 3.1, where 0+ and 1+ indicate the new logic values with the clock level shifted up. As a result, the final operation is $(A \rightarrow (D \rightarrow 0) \rightarrow 0$, where B is D $\rightarrow 0$. This operation is equivalent to A AND D, which is the required function to have the circuit operating in divide-by-3 mode. The divide-by-2 operation mode is performed by the using the normal clock DC level, where the circuit operates as a flip-flop connected in feedback mode. Consequently, a divider-by-2/3 is designed using just one flip-flop by considering the NIMP and NOT operations without requiring extra logic gates.

Experimental results indicate a 2.5 folds of speed improvement and a 4 folds of power reduction compared to the traditional AND and OR based frequency dividers presented in [17]. The measured output of the frequency prescaler in divide-by-3 operation mode, is presented in Fig. 3.7(a) given an input frequency of 12.3GHz. The prescaler is implemented in 0.13μm 1.2V CMOS technology [18]. Figure 3.7(b) shows a chip microphotograph. A photo of the measurement setup is presented in Fig. 3.7(c).

3.4 Using Flip-Flop+IMP and Flip-Flop+NIMP cells

Alternatively, using implication logic concept, it is possible to design the prescaler in an optimized manner. Figure 3.8 presents the dual-modulus prescaler implemented using the logic cells NIMP and IMP without requiring extra logic gates. As a result of the reduced complexity, low-power and high speed performance is achieved. Experimental results indicate an speed improvement of 75% and reduced power of 20% compared to traditional NAND and NOR based prescalers [17]. The measured

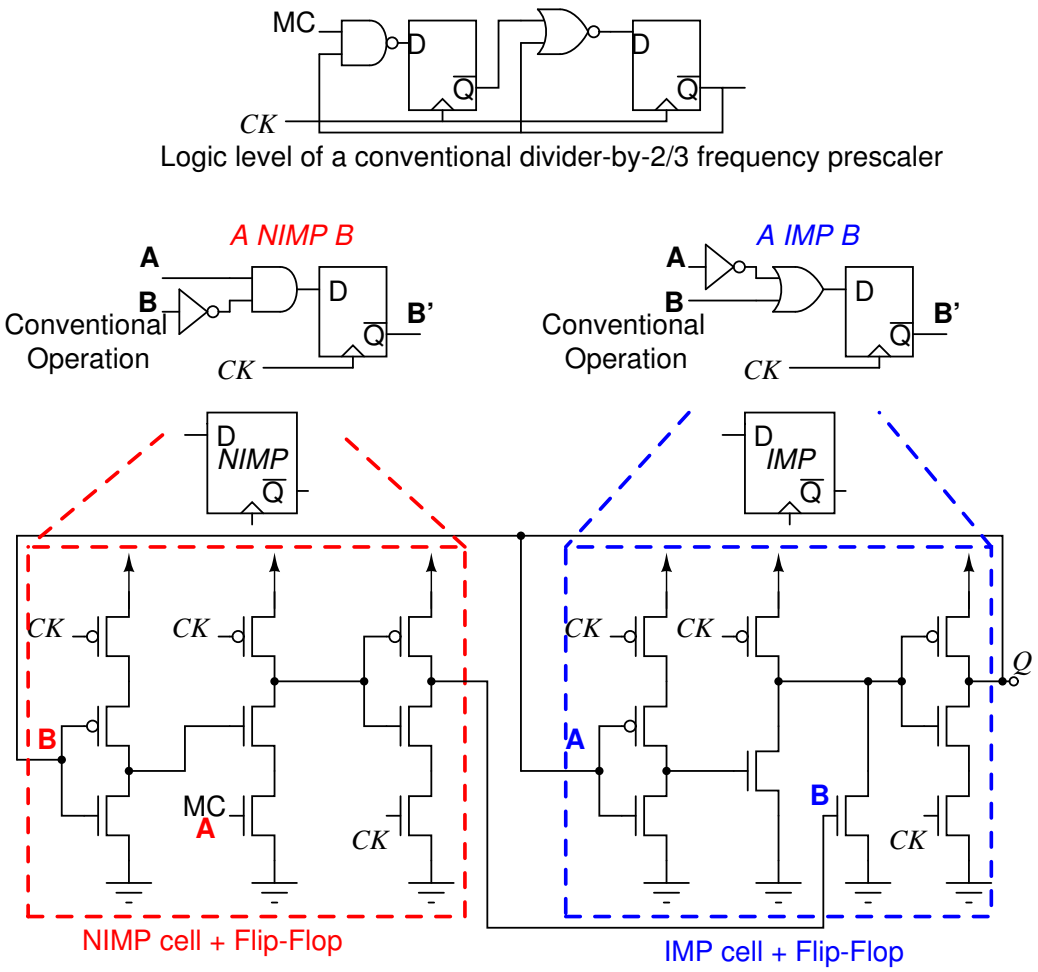


Fig. 3.8: Divider-by-2/3 frequency prescaler implemented with NIMP and IMP operations.

output of the frequency prescaler in divide-by-3 operation mode (MC=1), is presented in Figure 3.9 given an input frequency of 12.3GHz. The prescaler is implemented in 0.13μm 1.3V CMOS technology [19].

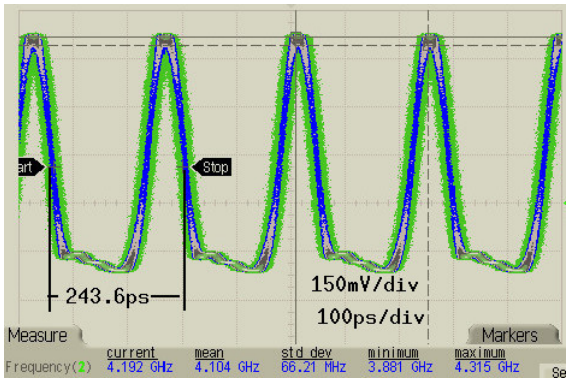


Fig. 3.9: Measured output in divide-by-3 operation mode.

3.5 Using Flip-Flop+IMP and Flip-Flop+NIMP cells with two transistors per branch

Figure 3.11 shows the proposed divide-by-2/3 FP, which removes logic gates in the critical path and turns-off the non-used branches, allowing a higher frequency operation with decreased power consumption. A block diagram of the proposed concept is shown in Fig. 3.10. In divide-by-3 mode, MC=1, the DFF2 output goes to a lower voltage level than V_{DD}, making M4s V_t lower. With a reduced V_t, M4 is stronger than the NMOS M3 in the ratioed logic. Consequently, b1 node is forced to high-level which makes DFF1 to hold for another cycle. A timing diagram describing the operation in divide-by-3 mode is shown in Fig. 3.12. The DFF2 output for the low-level is adjusted to a higher voltage in order to avoid M4s source-body junction of going fully forward-biased. This reduces the required driving capability at the output of DFF2, resulting in smaller transistor-size which relieves load capacitance and thereby, power consumption. In divide-by-2 mode, the output of DFF2 goes toV_{DD}, making DFF1 connection as a regular Mobius counter-by-2.

Using negative-edge DFFs instead of positive-edge ones decreases the load capacitance driven by the clock. In the case of positive-edge DFFs as in [3,19], Fig. 3.3, the clock is driving four gates of PMOS transistors. In the case of negative-edge DFFs

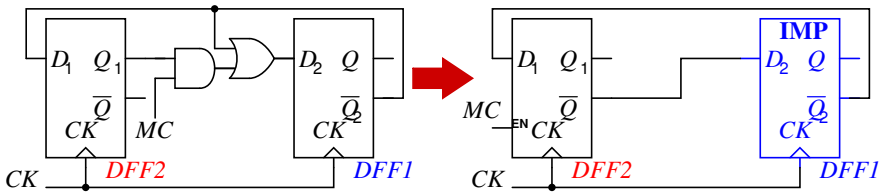


Fig. 3.10: AND-OR-based divider prescaler to implication-logic-based scheme.

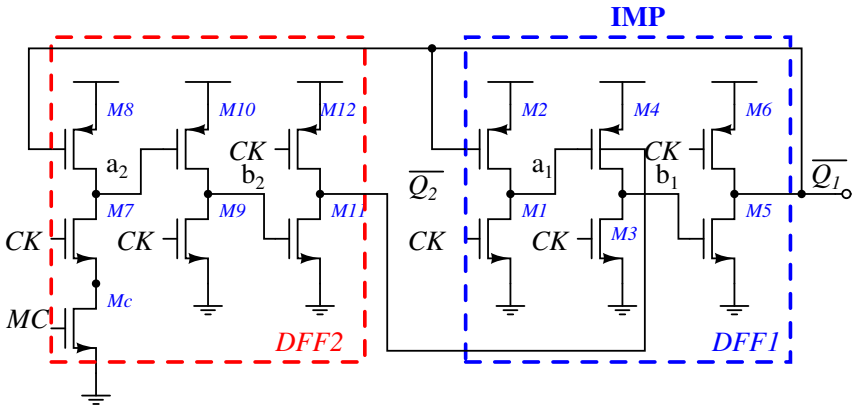


Fig. 3.11: Proposed two flip-flops (2FFs) dual-modulus prescaler.

Fig. 3.11, the clock is driving four gates of NMOS transistors. Therefore, the load capacitance to the clock is lower in the NMOS case considering that NMOS transistors are smaller for a given load, resulting in a power consumption reduction and a lower path delay. Moreover, the negative-edge DFF exhibits the characteristic that the branches are not switching if the first branch is disabled.

Instead of using the conventional logic concept, we rely on a new logic concept that allows us to minimize the number of elements in the critical path of the divider. The compact design is the result of using the implication logic concept [18]. In place of using the basic logic cells AND and OR to operate the selection mode, the DFF1 and DFF2 intrinsically execute the operations material implication IMP and its inverted operation NIMP, respectively. The truth tables of the IMP and NIMP operators are

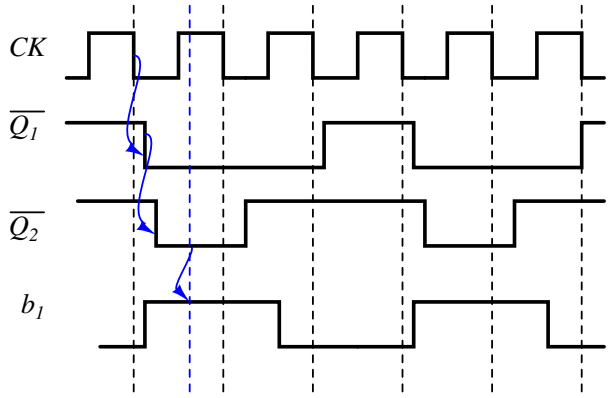


Fig. 3.12: Timing diagram in divide-by-3 operation mode (MC=1).

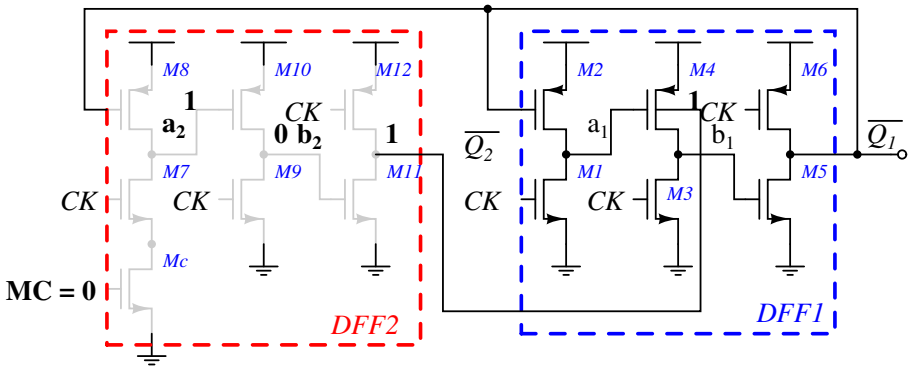


Fig. 3.13: Prescaler in divide-by-2 operation mode (MC=0).

also shown in Fig. 3.3. A latch or flip-flop that can execute the implication operation between its input and output would be able to synthesize Boolean logic operations. Recursive logic gates can be synthesized with implication operators that have the ability to store a variable. For instance, the FP in Fig. 3.11 shows that the DFF1 has the capacity to execute the IMP operation by replacing the OR operation in the classical divider-by-3. Considering the inputs of DFF1 as the gate of M2 and the body terminal of M4, and checking each combination of the inputs with the truth table of the IMP operation, the DFF1 executes the IMP operation.

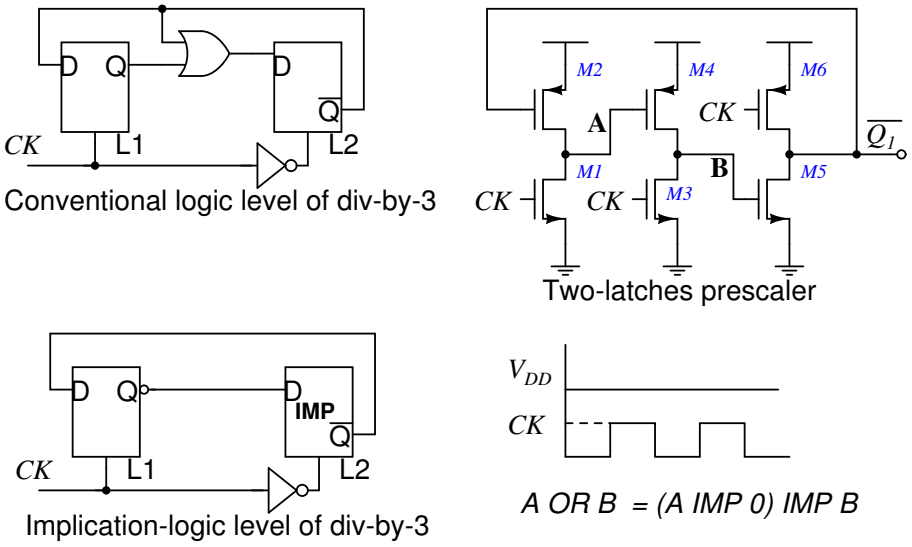


Fig. 3.14: Two latches dual-modulus prescaler using implication logic and IMP NIMP truth tables.

3.5.1 Proposed Two-Latches Prescaler

A simplified prescaler is shown in Fig. 3.14. In divide-by-2 mode, the FF is connected as a Mobius counter. The FF can be seen as two-latches, where the first latch is high-level active and the second latch is low-level active. By placing an OR gate between the two-latches, the circuit is setup as a divider-by-3. The operation OR is substituted by the operation IMP, where the second latch executes the IMP operation between the input and the output if the offset-level of the clock signal is shifted down. In other words, the operation performed at node B is (A IMP 0) IMP B, which is the same as A OR B. The complete performed operation is explained in Table 3.2, where 0- and 1- indicate the new logic values with the clock level shifted down. As a result, by shifting down the clock level, the circuit goes from divide-by-2 to divide-by-3 mode. This configurable behavior can leverage an automatic adjustment of closely spaced ratios.

Table 3.2. Result in node B for lower DC levels at CK.

A	B	$A_{CK=0}$	$A_{CK=1}$	$A_{CK=1-}$	$A_{CK=0-}$
0	0	1	0	1	1
0	1	1	0	1	1
1	0	0	0	0	0
1	1	1	0	1	1

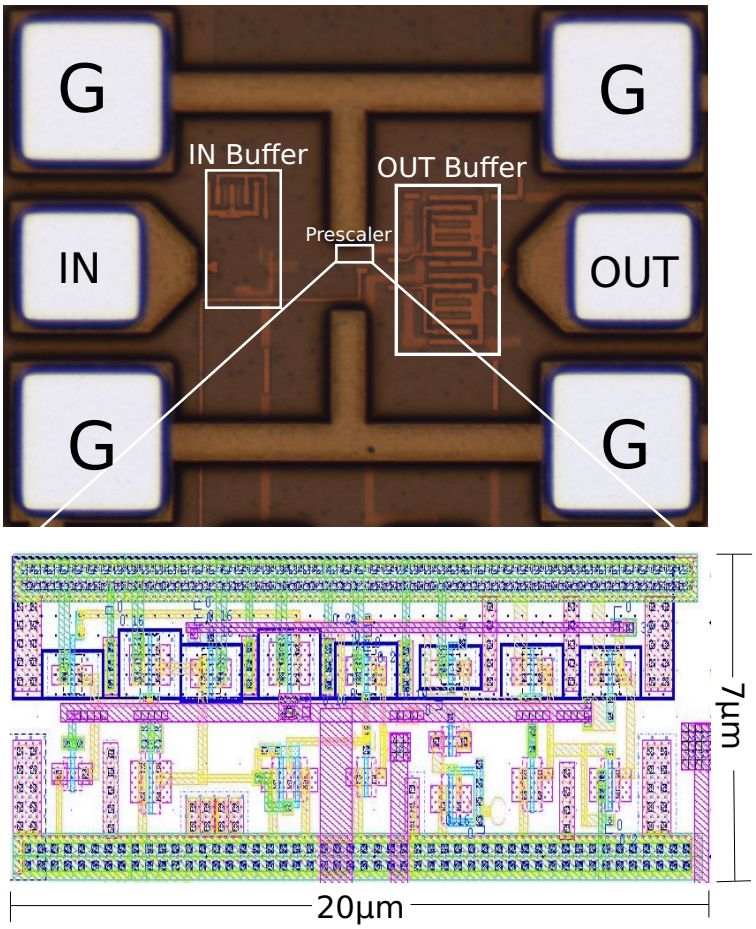


Fig. 3.15: Chip micrograph and layout details of the 2FFs prescaler.

3.6 Measurement Results

The dual-modulus prescalers were fabricated using 0.13μm CMOS and have an active area of 20μm $\times$ 7μm for the 2FFs based and 10μm $\times$ 7μm for the two latches

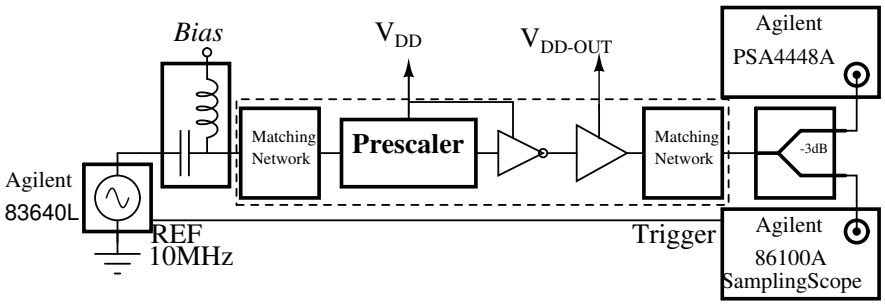


Fig. 3.16: Block diagram of the test setup.

Table 3.3. Performance comparison.

Design	**Tech.**	**Divisor**	f_{min}/f_{max} **[GHz]**	**Efficiency [GHz/mW]**
CML dynamic [20]	65nm	4	20/70	10.86
IL [1]	65nm	2	79/81.6	30.3
TSPC-based [2]	0.13μm	2/3	8GHz	6.8
TSPC-based [19]	0.13μm	2/3	5/14.1	11.66
TSPC-ratioed [3]	65nm	2/3	6.2/16GHz	35.71[!]
ETSPC [3]	65nm	2/3	8/18GHz	28.5
This work-IMP2FF	0.13μm	2/3	5.8/16.5	36.26
This work-IMP2L	0.13μm	2/3	6/17.1	36.3

[!] Measured in divide-by-2 mode.

based one. Figure 3.15 shows a micrograph of the test chip. The test setup is shown in Fig. 3.16. Notice that voltage supplies of the prescaler and output driver are separated to measure independently the power consumption. However, the power consumption of the first buffer seen by the output of the prescalers, is included in the reported consumption. For testing the two-latches prescaler, a bias tee was used at the input in order to shift the offset-level of the clock signal. Table 4.2 summarizes and compares the performance of the state-of-the-art FPs. The measured power consumption of the 2FF prescaler in divide-by-3 mode is 455μW from a 1.2V supply

at f_{max} of 16.5GHz, which is relatively larger than divide-by-2 mode, 339μW. The two-latches prescaler consumes only 471μW from a 1.2V supply at f_{max} of 17.1GHz in divide-by-2. In divide-by-3 mode, the consumption is 444μW.

Figure 3.17 shows the output waveforms measured with a sampling scope (Agilent 86100A) at f_{max} for the two-FFs FPs in the left-plots, and for the two-latches FPs in the right-plots. In Fig. 3.18, the mean measured input power level sensitivity plots for the 2FFs and two-latches (1FF) FPs are presented. The locking range at an input power of 12dBm is from 6GHz to 17.1GHz with a 1.2V supply for the two latches case. Figure 3.18 indicates an operation of 6GHz with a 0.55V supply voltage. Out of four measured chips, all are fully functional with variations of f_{max} in a range below 200MHz for different supply voltages and worst-case temperature. These results indicate the feasibility of using the FP in a wide frequency range with a low voltage supply settled to the required f_{max} to decrease power consumption.

Figure 3.19 shows the measured phase noise at the input (top-curve) for a 17.1GHz input. The middlemost-curve is the measured phase noise at the output in divide-by-2 mode. The phase noise in divide-by-2 mode is 6dB lower (up to 1MHz offset) than the input, as theoretically expected from a division-by-2. A similar result is obtained from the output phase noise in divide-by-3 mode (bottom-curve), which is 9dB lower compared to the input. The results indicate that the FPs introduce a negligible degradation. Measured spectrum plots at the output of the prescalers are shown in Figs. 3.20 and 3.21.

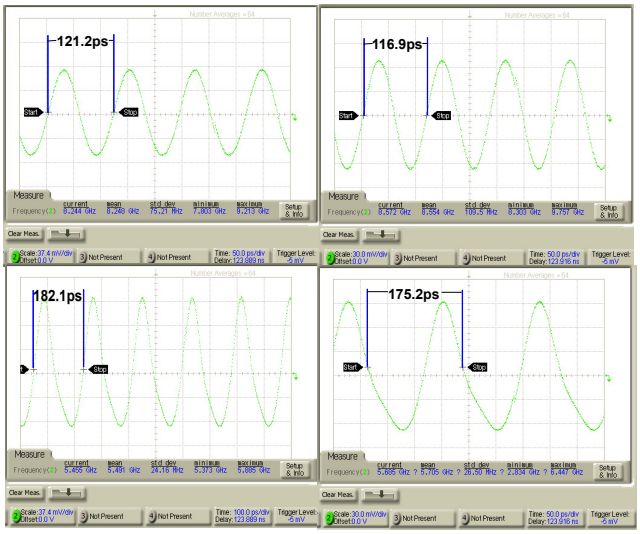


Fig. 3.17: Measured output waveforms in div.-by-2 mode (upper) and div.-by-3 (lower) for a 16.5GHz input in the 2 FF prescaler (left), and for a 17.1GHz in the two latches prescaler (right).

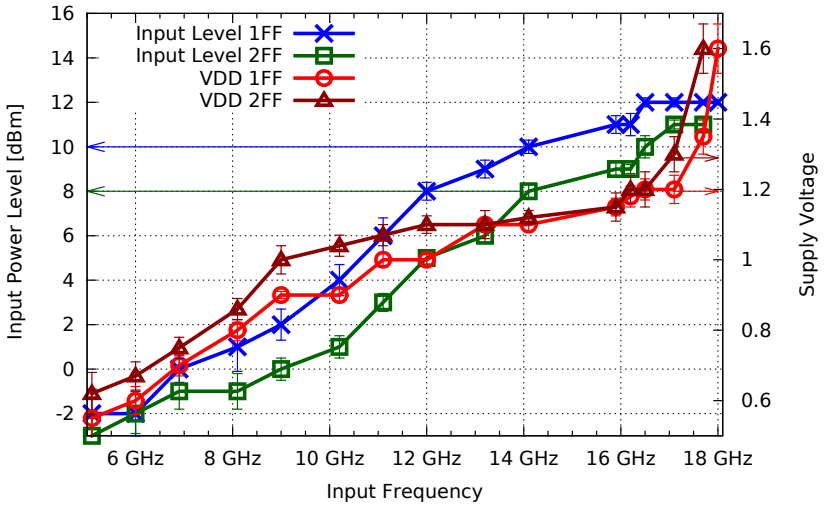


Fig. 3.18: Measured input power level and voltage supply sensitivity in divide-by-3 operation mode.

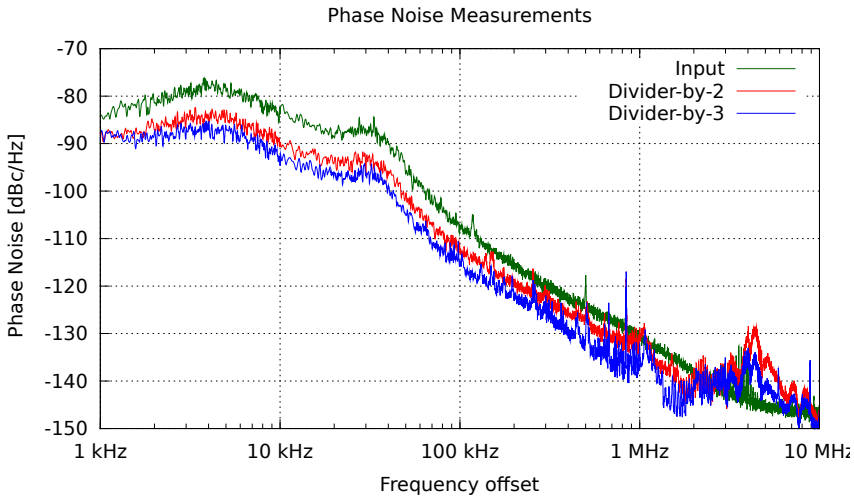


Fig. 3.19: Measured phase noise at the input (top curve), at the output in divide-by-2 mode (middlemost curve) and -3 operation mode (bottom curve).

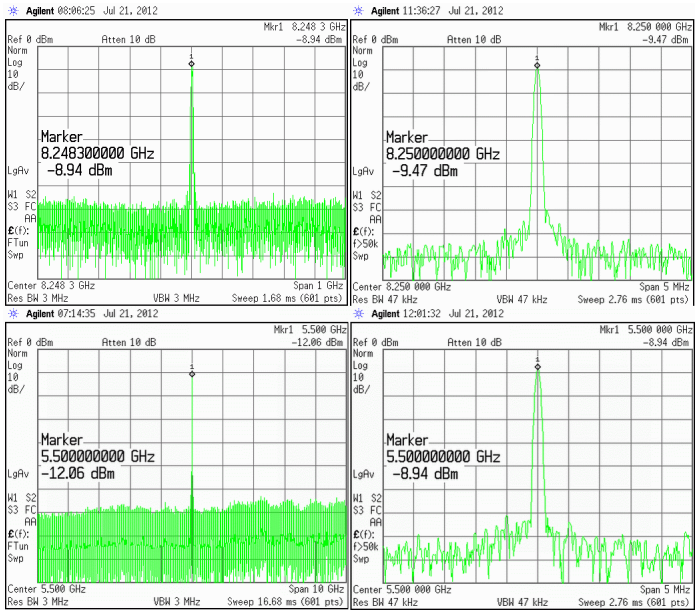


Fig. 3.20: Measured spectrum output in divide-by-2 mode (top plots), and in divide-by-3 mode (bottom plots) at 16.5GHz input for the 2FF prescaler.

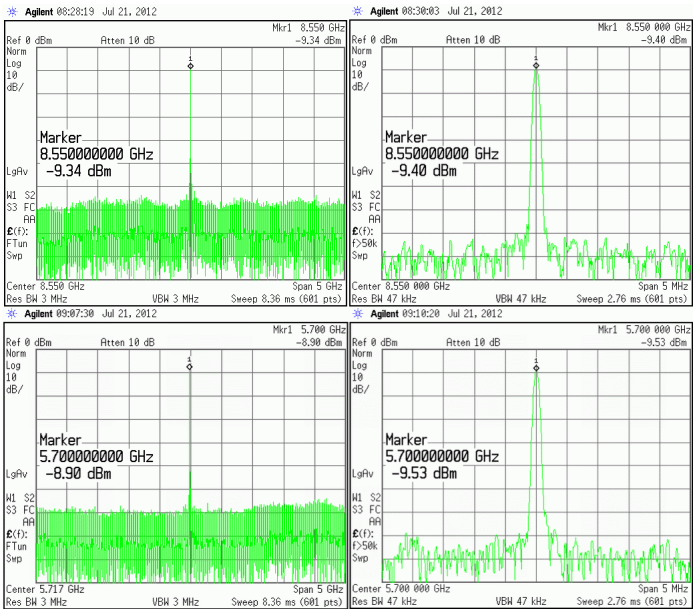


Fig. 3.21: Measured spectrum output in divide-by-2 mode (top plots), and in divide-by-3 mode (bottom plots) at 17.1GHz input for the two-latches prescaler.

4. WIRELINE TRANSMITTER

The accelerated bandwidth requirement for high-speed interconnections in mobile and server applications, demands improved link power efficiency and increased serial data rates at lower cost. In a serial link transceiver, the multiplexer (MUX) plays a key role in the maximum achievable data rate and constitutes a significant portion of the overall power consumption. The tree topology is the most popular MUX architecture due to its ability to operate at high speeds. Although 4:1 MUX tree topology, Fig. 4.1(a), might require a higher number of gates at the datapath compared to a single-stage MUX, as the shown in Fig. 4.1(b), single-stage multiplexers have a large parasitic loading at the output node and require multi-phase clock generators degrading the speed and power efficiency performance.

Conventional high-speed MUX operating at +10Gb/s are implemented using current mode logic (CML) [21, 22]. A differential CML flip-flop (FF) is capable of faster operation compared to other CMOS logic in a given technology node. CML is faster due to its current-steering operation in small signal domain. However, the power consumption is determined by the sum of the two saturation drain-currents in the differential pair. Moreover, a relatively large supply voltage is required in order to have sufficient source-drain voltage to guarantee operation in the saturation region for each transistor. In contrast to CML, CMOS logic power consumption depends on the operation frequency and is usually lower than CML, even at high speeds. Recently, CML has been adopted in the output stage to achieve high throughput, and CMOS static logic is now employed in aggressive manner at the medium and low-speed stages [21, 23].

This work presents a new scheme for high-speed MUX using CMOS-based logic. We have demonstrated a low-area 4:1 MUX operating at 40Gb/s based on true-single-clock-phase (TSPC) logic-style. A power efficiency of 21.5μW/Gb/s or 21.5fJ/b is

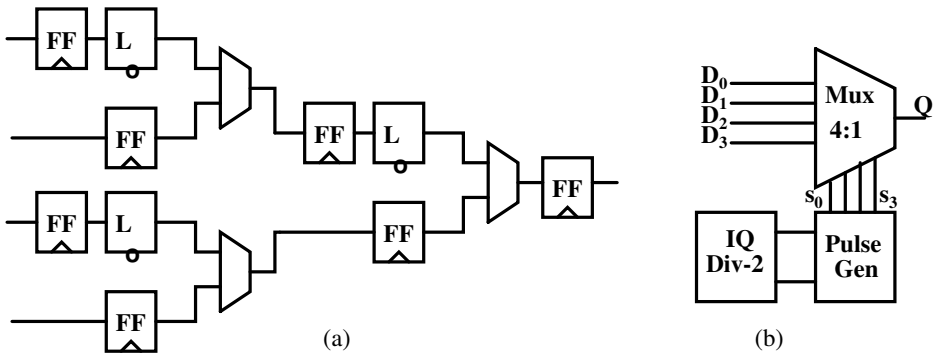


Fig. 4.1: 4:1 multiplexer (a) conventional tree topology, (b) one-stage multiphase clock.

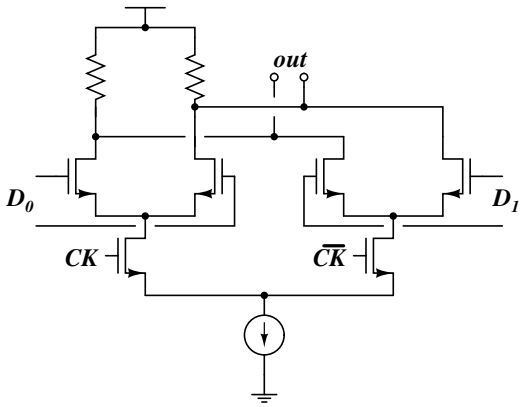


Fig. 4.2: CML-based 2:1 multiplexer.

achieved in 45nm SOI CMOS technology. Low power consumption is obtained by the use of TSPC circuits in the reduced-speed stages, and extended-TSPC (ETSPC) circuits in the output stage. Clock distribution is minimized by using FFs with single-phase clock operation, thus power consumption is decreased considerably in the high-speed stages. Furthermore, we propose a MUX-FF-Latch (MFFL) combination based on ETSPC logic capable of re-timing, time shifting and selection functions like the FFs, latch and MUX in the conventional tree scheme. As a result, a reduced number of gates enables low-power operation at +20Gb/s.

4.1 4:1 Multiplexer Design

Conventional tree topology uses FFs to re-time the data streams in order to align input data with the clock signal. In addition, input data at the 2:1 MUX must be offset in half clock period by placing a latch to avoid simultaneous transition of the MUX inputs. The employment of CML is the common approach to design high-speed 2:1 MUX and latches, considering that data rates of static CMOS FFs are limited by the toggle frequency. However, using TSPC logic-style, data rates can be improved significantly. Figure 4.4 shows the tree topology employed in the 4:1 MUX design., denoting the TSPC usage in the front blocks. Recently, TSPC FFs have been used in high-speed prescalers with high power efficiency [18, 19].

One key advantage of using TSPC logic-based is the not need for a differential clock generation as in CML, therefore, skew/overlapping effects do not take place which reduces jitter introduced in the output data. Additionally, clock path distribution is reduced which eliminates the need for extra high-speed clock buffers. The 2:1 MUX at the output, uses ETSPC logic considering the reduced load in ETSPC cells and consequently faster operation. TSPC circuits are used in the low-speed stages where the clock sinks are driven by half frequency of the clock source, as indicated in Fig. 4.3. To describe operation of the topology used (Fig. 4.3(a)), a timing chart is shown in figure 4.4. An exploded view of the architecture is presented in Fig. 4.3(b), indicating the signal nodes used in the timing diagram of Fig. 4.4.

4.1.1 ETSPC 2:1 MUX

The proposed high throughput 2:1 MUX is presented in Fig. 4.5. The MUX uses a positive-edge FF and a latch transparent in high-level (MUX-FFL), implemented with ETSPC logic-style. While the latch is in the evaluation phase, the FF is in the holding phase, and vice versa.

Despite the fact that ETSPC is a ratioed logic, transistors are sized such that at high frequency operation the short current period occurs at the same period of

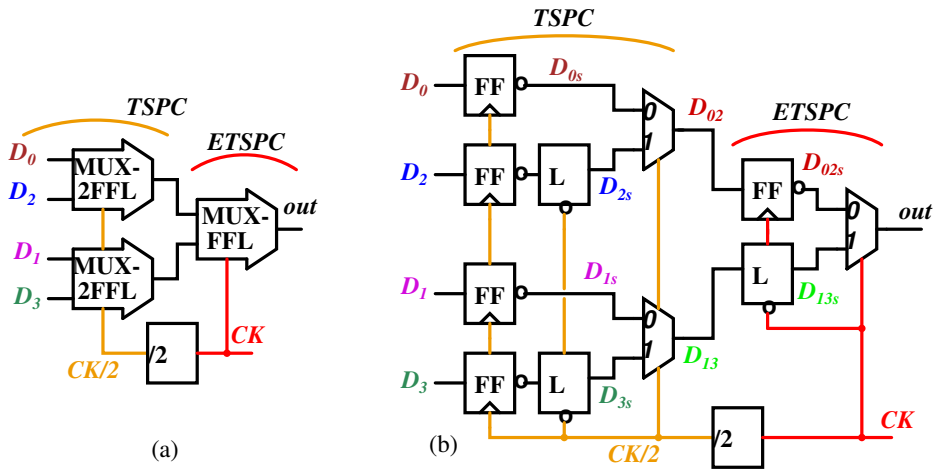


Fig. 4.3: Architecture of 4:1 multiplexer: (a) with MUX-FFL blocks and, (b) exploded block schematic.

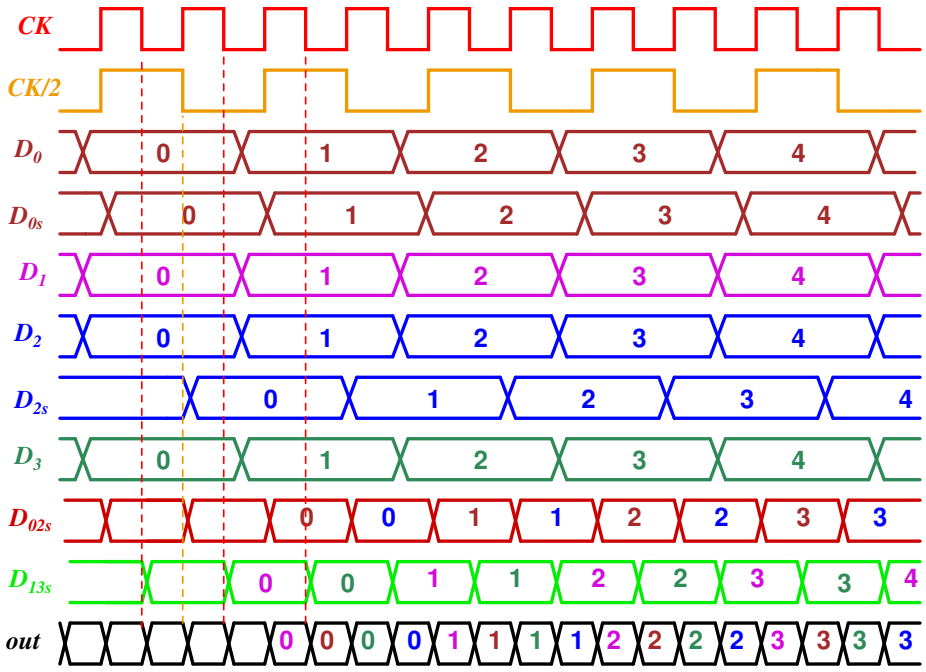


Fig. 4.4: Timing diagram of the 4:1 multiplexer.

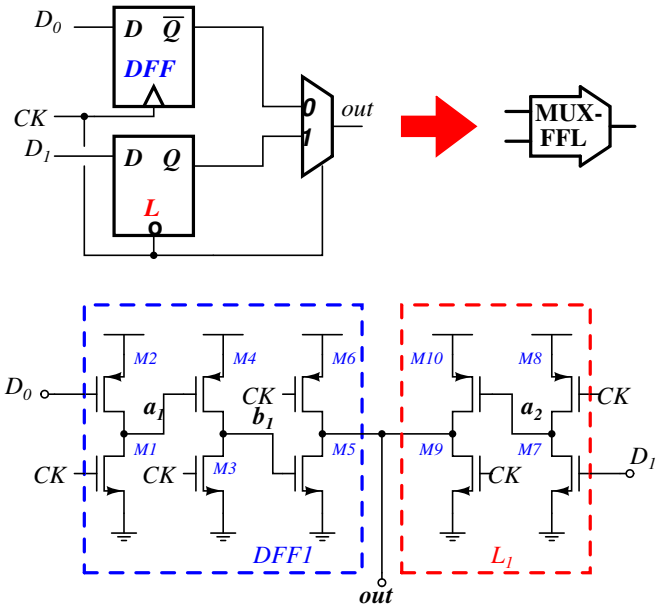


Fig. 4.5: A combination of FF, latch and 2:1 multiplexer proposed using ETSPC logic-style.

switching level. As a result, an energy efficient operation is achieved. In the latch configuration, M8 is stronger than M7, and M10 is stronger than M9. Here, when CK is low, the output of L1 in high impedance, and the a2 node holds the inverted level of D1. When CK goes high, M9-M10 evaluates the value at a2 node. At this time, if a2 is high-level, M10 turns off and the output goes low. If a2 is low-level, M10 turns on dominating over M9 and the output goes high.

This configuration is able to perform re-timing, timing shift and selection functions at the same time. In contrast, a traditional 2:1 MUX realizes the functions separately, where two FFs, a latch and selector are required.

4.1.2 TSPC 2:1 MUX

At lower frequency operation, TSPC cells are more energy efficient than ETSPC ones. TSPC-based 2:1 MUX are used at the input data of the 4:1 MUX. Considering

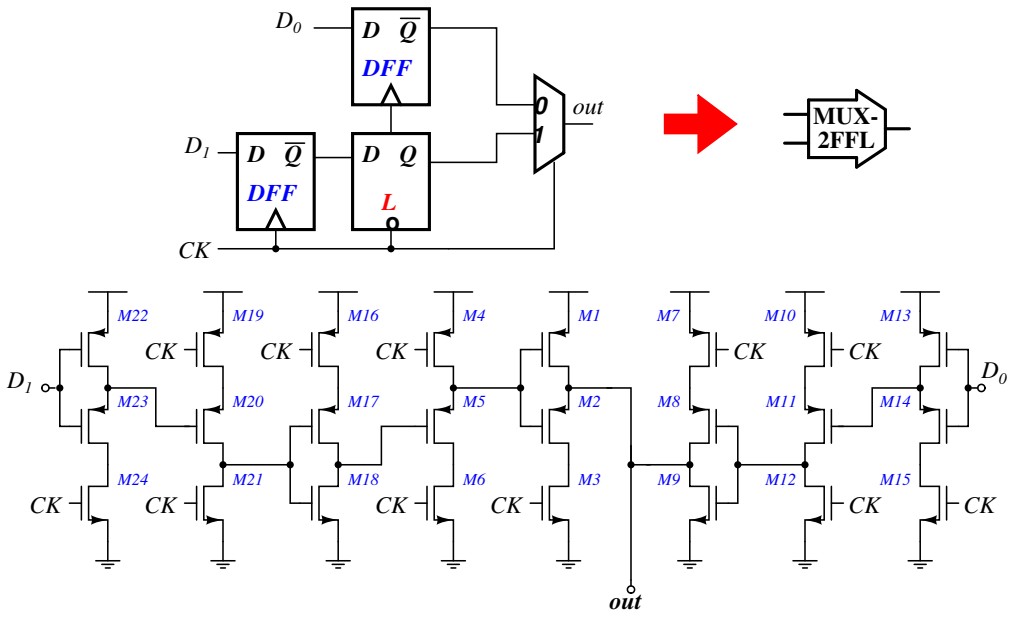


Fig. 4.6: A combination of FF, latch and 2:1 multiplexer proposed using TSPC.

the possibility of misalignment of the input data-source, two FFs are used to align the data to the clock signal. Figure 4.6 depicts the configuration and transistor-level schematic of the MUX, FFs, and latch combination (MUX-2FFL). Similar to the MUX-FFL, the FFs are triggered at the positive edge and the latch is transparent in high-level. Despite the higher number of stacked transistors and as a consequence higher load capacitances, there is a reduced racing between the PMOS transistor connected to V_{DD} and the NMOS butted to ground [13].

4.2 Pseudorandom binary sequences generator

Reliable testing at high speed data rates require pseudorandom binary sequences (PRBS). Performance of high data rate must be evaluated under stressing signals that accumulate a long series of high/low digital logic values. However, the high cost of high data rate commercial PRBS generators can be impractical if multiple input signals are required. Furthermore, the I/O number of pins to test a serializer can

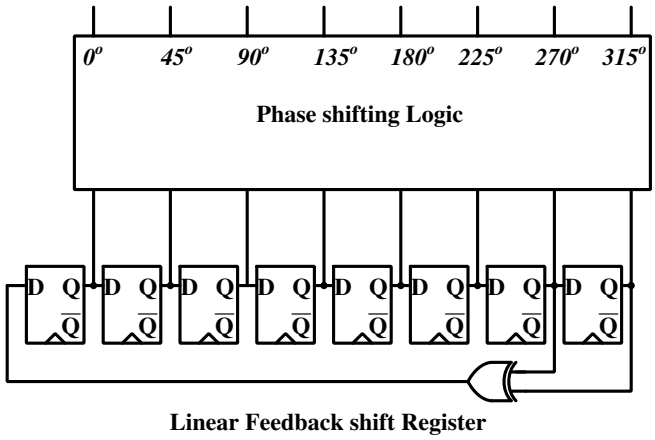


Fig. 4.7: Traditional $2^7 - 1$ PRBS generator using LFSR and phase-shifting logic.

impact the overall chip area which translates to a higher silicon cost. Therefore, a build-in low area PRBS generator is preferred. Although, the power consumption of the PRBS generator might not be relevant during normal link operation, it might be appropriated to have low power dissipation in case the link application require self-testing and auto-calibration modes.

Prior art of PRBS generators use extensively linear feedback shift registers (LFSR) due to the straightforward implementation [24]. LFSR-based PRBS generators require large number of gates with large fanouts that impact in the maximum achievable speed. In addition, if multiplexing techniques are used to increase the output data rate, additional phase shifting techniques are required, which results in an increased complexity and power dissipation [25]. Figure 4.7 shows a conventional LFSR-based PRBS generator with a phase shifting logic block to produce further multiplexing.

An alternative to LFSR-based generators is to employ parallel PRBS generation, that uses a reduced number of gates leading to a higher achievable frequency operation [25, 26]. Moreover, parallel PRBS generators are suitable to further output multiplexing since their outputs are intrinsically shifted. In [25], Laskin et. al report a high data rate parallel PRBS generator using BiCMOS CML latches. However,

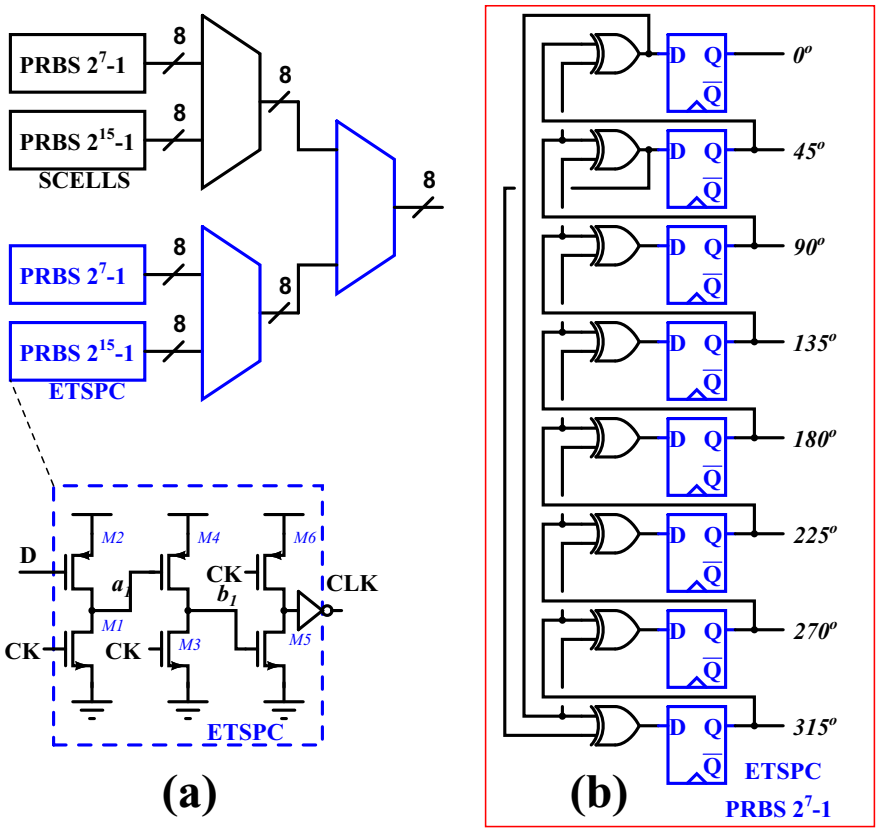


Fig. 4.8: Proposed PRBS parallel generator using high-speed ETSPC flops.

CML latches impact the power budget since they are power hungry circuits. In this work, a parallel PRBS generator capable of achieving 10Gb/s data rate before output multiplexing is presented. Instead of using CML latches, ETSPC latches are employed. Exploiting the high-efficiency characteristic of ETSPC latches, low area and low power PRBS generators are implemented. One generator with a $2^7 - 1$ length sequence and the other one with $2^{31} - 1$ length sequence, Fig. 4.8. Two PRBS generators using static flip-flops were also implemented for comparison purposes. Figure 4.14 shows the layout of the $2^7 - 1$ proposed PRBS generator which occupies an area of $80\mu m^2$. The PRBS generator design is suitable for a standard-cell flow. The layout was fully placed and routed using a commercial flow which generates a more compact placement.

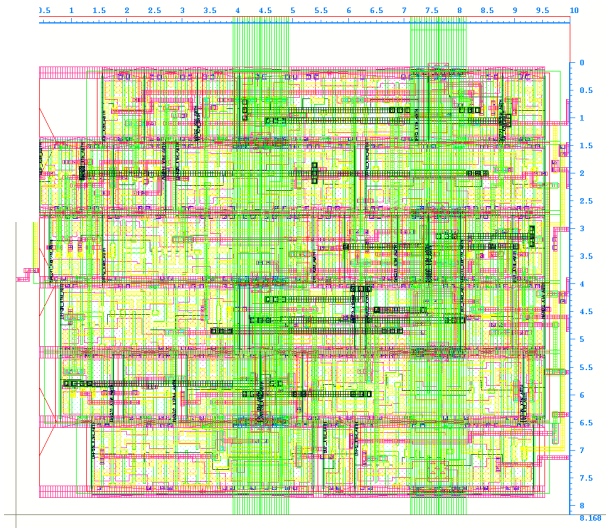


Fig. 4.9: Fully placed and routed $2^7 - 1$ proposed PRBS generator.

Alike the PRBS generator, it is also desirable to have a build-in PRBS checker to measure the link performance. Figure 4.10 shows the proposed PRBS scheme which does not require a synchronization circuit. The floating XOR and DFF gates in the figure represent the mirrored PRBS generator placed at the receiver side. The proposed self-synchronized PRBS checker includes an error counter at its output to easily read the bit error rate. The output of the counter can be stored in one of two SPI registers without additional circuit complexity.

4.3 Simulation results

The full-rate 4:1 MUX implemented is shown in Fig. 4.11(a) with the re-timer FF and the divider-by-2 shown in Figs. 4.11(b) and 4.11(c) respectively. Pos-layout simulations performed in spectre-rf with nominal corner models, demonstrates operation at 40Gb/s. Monte-Carlo simulations across temperature up to 100°C indicated that the sizing of transistors is adequate to severe PVT variations. Slow-slow (SS) corner limited the input clock source to 44GHz from a 1V supply. Four different 10Gb/s

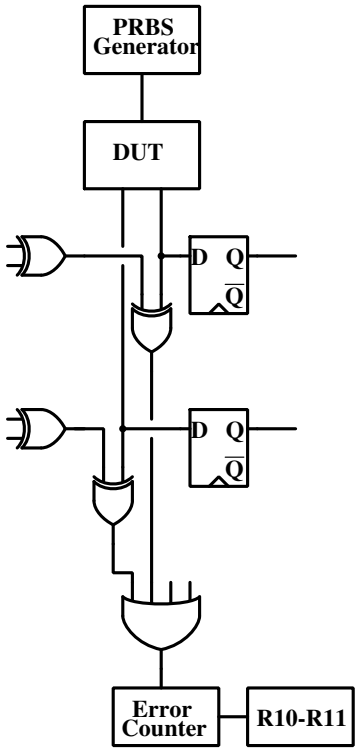


Fig. 4.10: Proposed PRBS checker scheme.

PRBS at the input were used in simulations with a clock source at 40GHz. Figure 4.12 shows the maximum input data rate obtained in the worst-case, and the respective power consumption as functions of the supply voltage. Given that ETSPC employs two stacked transistors between supply rails, its is possible to operate at low voltage for different input data rates. Consequently, power consumption can be decreased for lower data rates enabling power managing features at different speeds. For instance, at 24Gb/s the voltage supply can be as low as 0.7V with a power consumption of 303μW.

As comparison scale, the performance of reported 2:1 multiplexers are compared in Table 4.2. High-speed data rates (DR) are achieved with CML using inductive peaking technique [22, 27, 28]. However, inductor footprints increase the area significantly with area penalties over 0.6mm^2. Along with large area, reported power efficiency

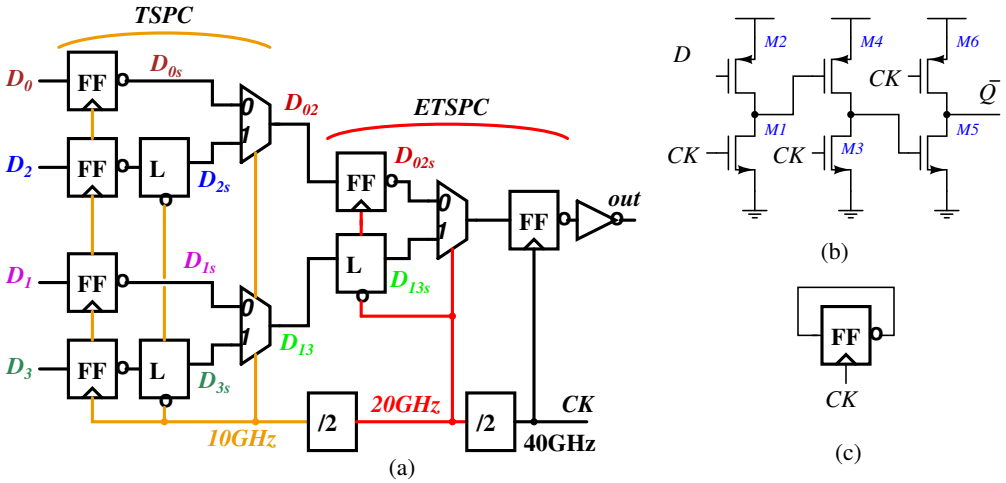


Fig. 4.11: (a) 4:1 MUX implemented. (b) ETSPC-based FF schematic. (c) Divider-by-2 scheme.

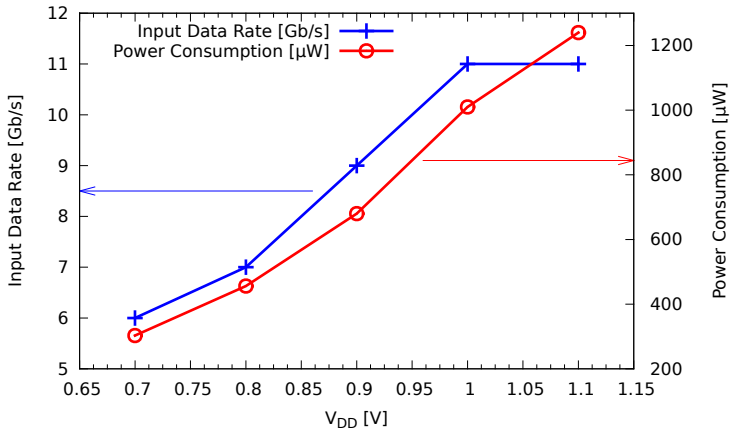


Fig. 4.12: Sensitivity of input data rate and power consumption to supply voltage.

is larger than 2mW/Gb/s. Hitachi reported in [23], a complete transceiver in 65nm CMOS with low power consumption. Although the energy efficiency reported in [23] (0.2mW) does not include the divider and re-timing circuit, the result highlights an energy-efficient performance. In 2011, Hitachi presented a 25Gb/s transceiver [21]. Besides the power consumption of the isolated MUX 2:1 is not reported, a power

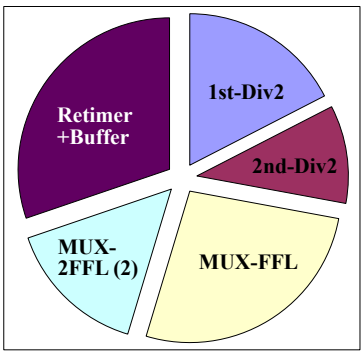


Fig. 4.13: Power consumption breakout out of 860μW at 40Gb/s of the 4:1 MUX implemented.

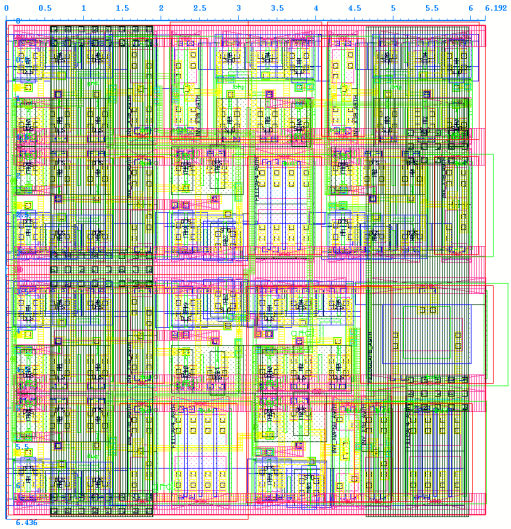


Fig. 4.14: Fully placed and routed high speed 4:1 MUX using PnR flow.

breakout of the transmitter indicates that the energy efficiency is not better than reported in [23]. For comparison, the energy efficiency of this work indicated in the Table 4.2 is just for the MUX-FFL cell (230μW/40Gb/s).

Energy-efficiency parameter suggests that the proposed MUX is the most efficient multiplexer operating above 10Gb/s reported to date. The 4:1 MUX occupies 20μm x 28μm of core area. A breakout pie chart is shown in Fig. 4.13 with the power distribu-

Table 4.1. Performance comparison of 2:1 Multiplexers.

Design	Tech.	Supply [V]	DR [Gb/s]	Energy Eff.
CML-Ind. [22]	90nm	1.2	40	270fJ/b
CML-Ind. [27]	180nm	1.2	40	14.7pJ/b
CML-1stage [28]	130nm	1.5	30	~1.1pJ/b
CML [29]	90nm	1.2	12	NA
CML [23]	65nm	1	12.5	16fJ/b
CML [21]	65nm	1	25	NA
This work	45nm	1	40	**6fJ/b**

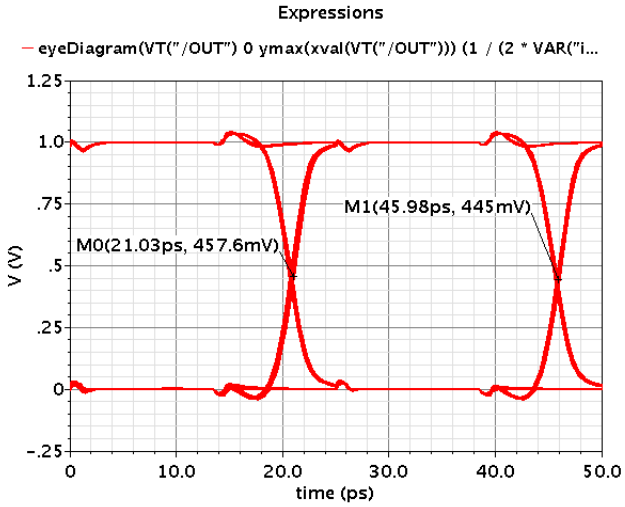


Fig. 4.15: 4:1 MUX eye diagram @ 40Gb/s after output re-timing.

tion of each block for the complete full-rate 4:1 MUX. The total power consumption of the 4:1 MUX at 40Gb/s output data rate is 860μW.

Figure 4.15 shows the eye diagram at the output of the 4:1 MUX. The eye diagram has 24ps horizontal eye opening at the slow-fast corner. Figure 4.16 shows the eye diagram at the output of the 4:1 MUX before the re-timer flop, with the transistors in D_{02} datapath (Fig. 4.11(a)) using slow-fast corner models, and the D_{03} datapath using fast-slow models. Further analysis suggests that it is possible to re-time the

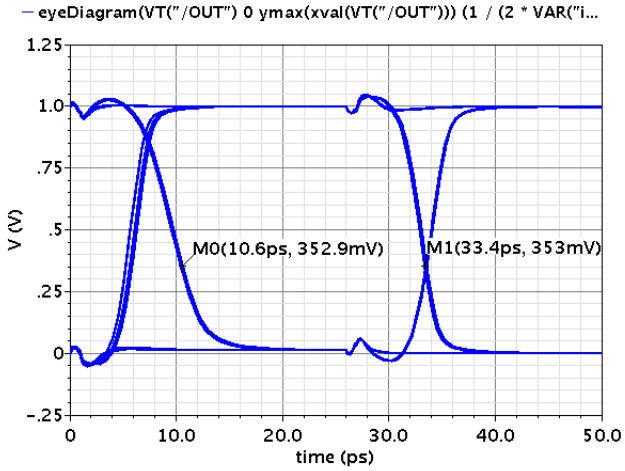


Fig. 4.16: 4:1 MUX eye diagram @ 40Gb/s before output re-timing.

output in the transmitter driver opening the possibilities to increase the limited speed by the full-rate MUX. For instance, a two-phase +40GHz clock signals with phase difference of 90° might be used to re-time the MUX output [21]. A sample of the 4:1 MUX output waveform at 40Gb/s output data is presented in Fig. 4.17. A pattern of 10Gb/s $2^{31} - 1$ PRBS was applied to each input.

Statistical analysis where run using Monte Carlo with 100 samples. Figure 4.18 presents the eye diagrams measured at the output of the driver for the most relevant Monte Carlo cases . Twenty cases are presented which include the corners: slow-slow, slow-fast, fast-slow at VDD=0.95V. Results indicate a 22ps horizontal eye opening and 280mV vertical opening. The jitter was measured at the typical corner considering the power supply induced jitter, Fig. 4.19. With a 10ps switching supply noise, the jitter measured at the driver's output is less then 3ps. The eye diagram at the end of the channel termination with the FFE off is presented in Fig. 4.20. Attenuated signal and ISI are observable leaving to a closed eye diagram. The eye diagram for different corners is shown in figure 4.21 with the FFE turned on indicates a 20ps horizontal and 148mV vertical eye opening.

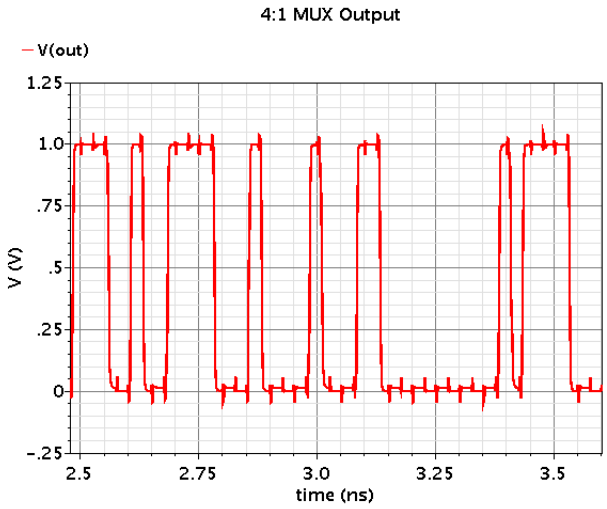


Fig. 4.17: 4:1 MUX output with input data rate of 10Gb/s.

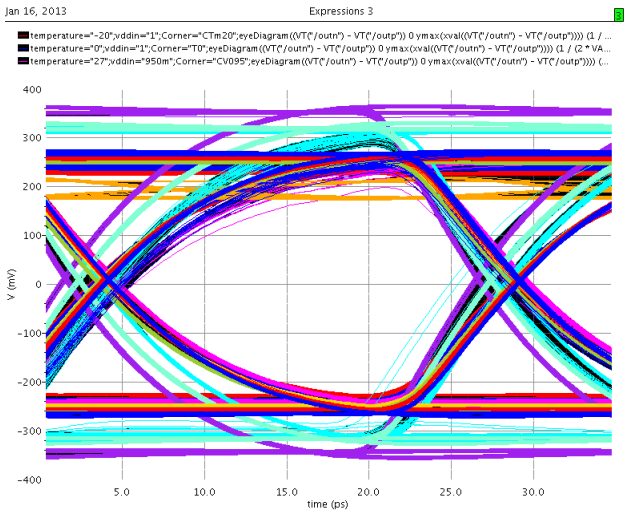


Fig. 4.18: Monte Carlo results of 20 relevant cases showing eye diagram @ 40Gb/s at the driver's output.

4.4 Implemented Transmitter

Fig. 4.22(a) shows the proposed TX that consists of PRBS generators, an 8-to-1 serializer, a 4-tap pre-emphasis equalizer, a source-series driver stage, and a deseri-

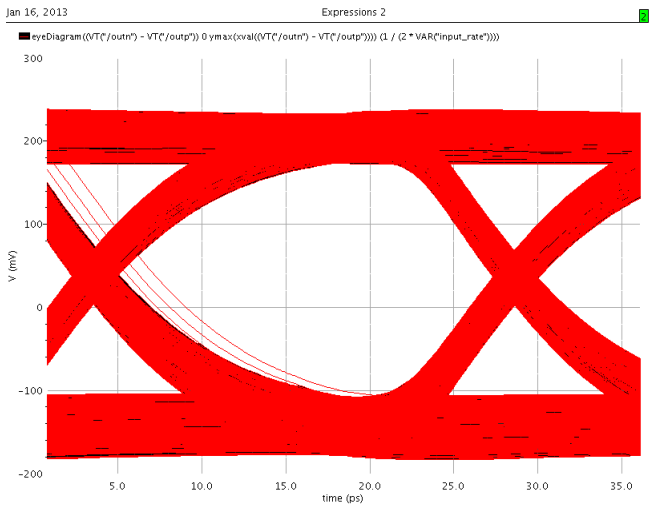


Fig. 4.19: Jitter measurement at typical corner @ 40Gb/s.

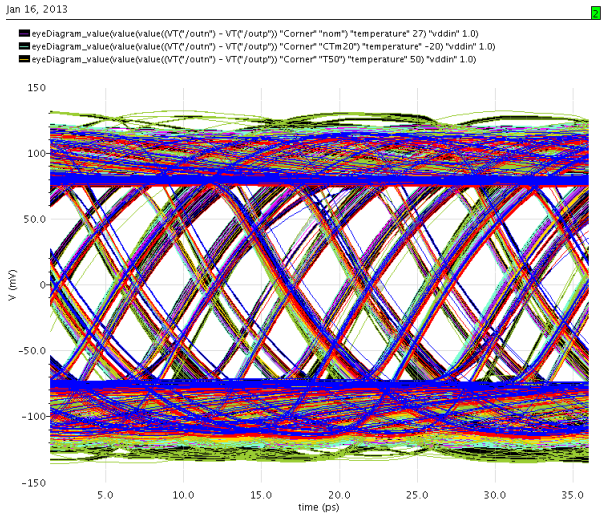


Fig. 4.20: Output at the 50 Ohm termination with FFE off.

alizer that includes a PRBS checker. In a common architecture, the pre-emphasis stage drives the output stage directly at full-rate or half-rate. Therefore, a reduced number of pre-emphasis taps is preferred in order to save power. However, in high loss channels a small number of taps might be not enough to equalize the channel.

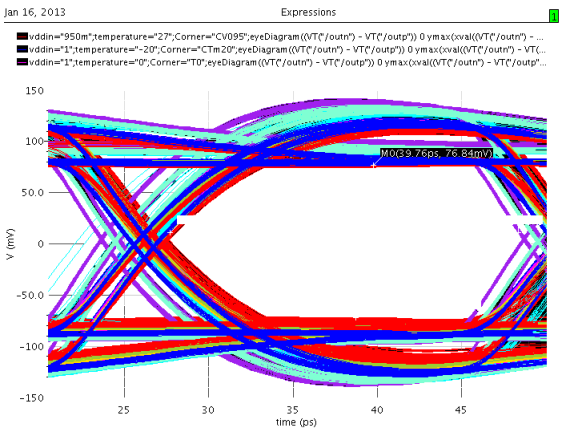


Fig. 4.21: Equalized output with FFE on at 50 Ohm termination.

The proposed TX has a half-rate architecture with the final 2:1 muxing stage after the pre-emphasis. Unlike the design in [30], the full-rate muxing of the data stream occurs before the driver which results in a significant reduction in the clock load. A pre-emphasis configuration works at quarter-rate to minimize power consumption of the digital logic. After the pre-emphasis, S2D blocks drive the final 2:1 mux with differential signaling to decrease power supply noise.

Each tap has a configurable driver segment with five weighted final drivers. Fig. 4.22(b) shows one driver segment that includes the last 2:1 mux and the selection gate of the final driver. Each pull-up and pull-down device of the driver has its own bit control to set the SST impedance and control the swing accordingly. With 118 unity cells in a 5-bit tunable pre-emphasis weight, 86 units are dedicated to pre-emphasis control with a resolution of about 2%. The final output impedance is tuned-up with 32 unit cells per driver segment. The pre-emphasis taps can be independently selected as pre-cursors and/or post-cursors providing a reconfiguration capability for different channel characteristics. Each segment includes a pre-driver data selection circuit that adjusts tap coefficients. As in [30], the SST driver does not employ stacking transistors in the pull-up and pull-down branches, eliminating the parasitic and therefore the stored data-dependent charges.

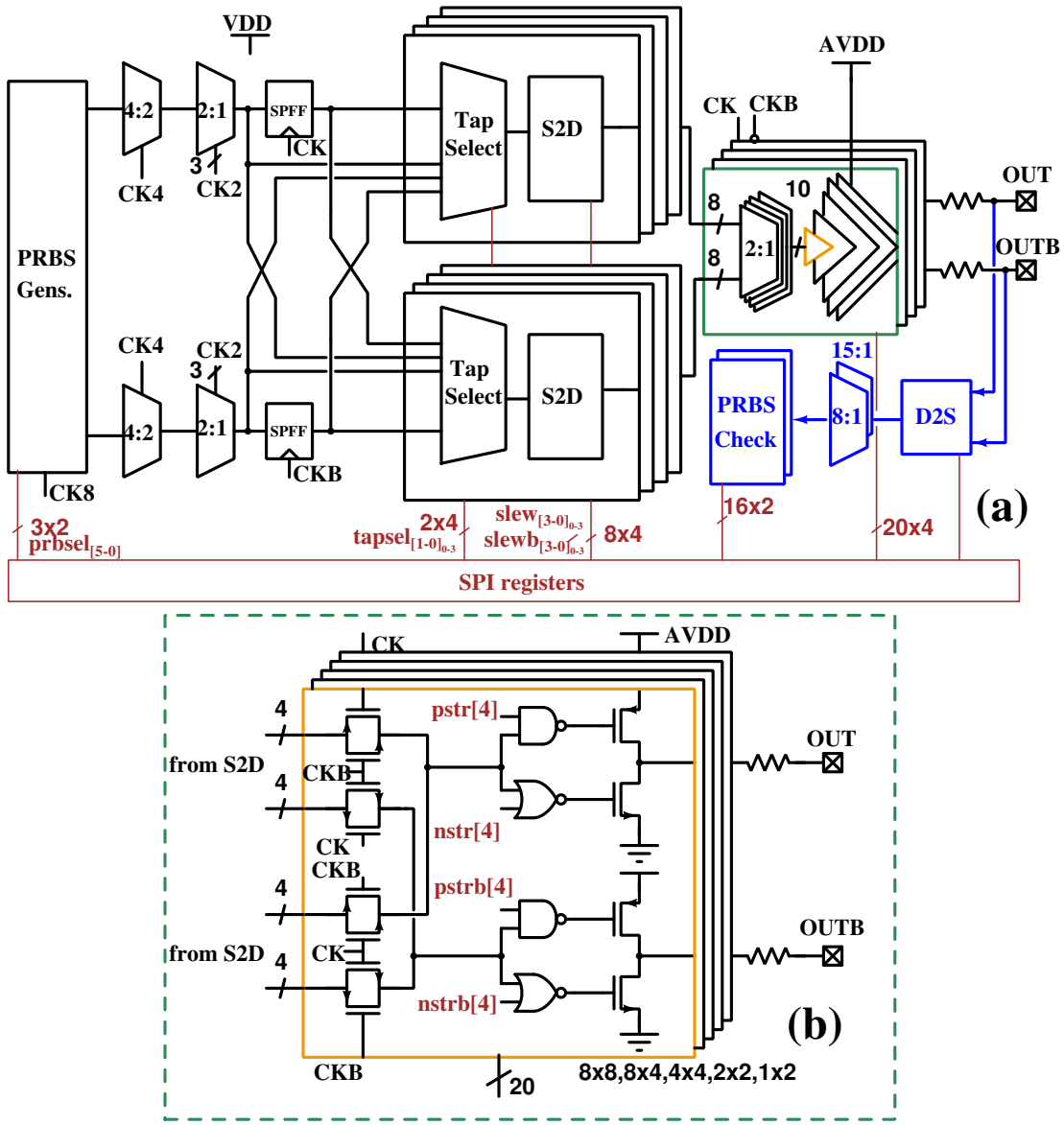


Fig. 4.22: (a) SST proposed TX architecture and (b) 1-tap driver segment with weight equalization control gates.

$2^{15}-1$ and $2^{7}-1$ parallel PRBS generators shown in Fig. 4.8 are implemented using standard cells and extended true single-phase clock (ETSPC) flip-flops. A PRBS checker is implemented using the same parallel PRBS scheme of the generator as shown in Fig. 4.10. At the checker, XOR outputs of the PRBS do not set new

states but are compared with the incoming stream of bits to flag errors if bits are not equal. This technique, called multiplication by a reciprocal polynomial [31] allows self synchronization to the incoming stream of bits without a synchronization circuit.

4.5 Transmitter Circuits

Fig. 4.23(a) shows the clock distribution circuit. A single-ended input clock is used to decrease power consumption for clock distribution, and a digital single-to-differential (S2D) circuit is placed close to the circuit that requires complementary clock signals. The AC coupling distribution decreases power consumption since the signal swing amplitude in the clock lines can be reduced. Skew control lines are independently implemented for the quarter-rate and half-rate clocks to alleviate transition mismatch issues.

4.5.1 Duty Cycle Correction

A duty cycle correction circuit (DCC) is proposed to adjust the duty cycles of the half-rate clocks CK and CKB. In addition to have delay line control in each clock path, S2D blocks and DCC are inserted to provide a programmable slew, and skew and duty control of the half-rate clock signals at the point of load. Figs. 4.23(b),(c) shows the proposed DCC circuit and its respective timing diagram. Considering that S2D differential block has a delayed branch due to the inserted inverter to generate the complementary signal, Fig. 4.24, the output of the S2D might generate a distorted clock signals (CKi, CKBi) with non-50% duty cycles. Assuming that CKi and CKBi signals have duty cycles greater than 50% as depicted in Fig. 4.23(c) with rising edges separated by half-period, the output clock Q will have a 50% duty cycle. For instance, assuming that the output Q is selecting the CKBi input at the output of the mux O_{MUX}, a rising edge of CKBi will trigger the FF therefore sampling an

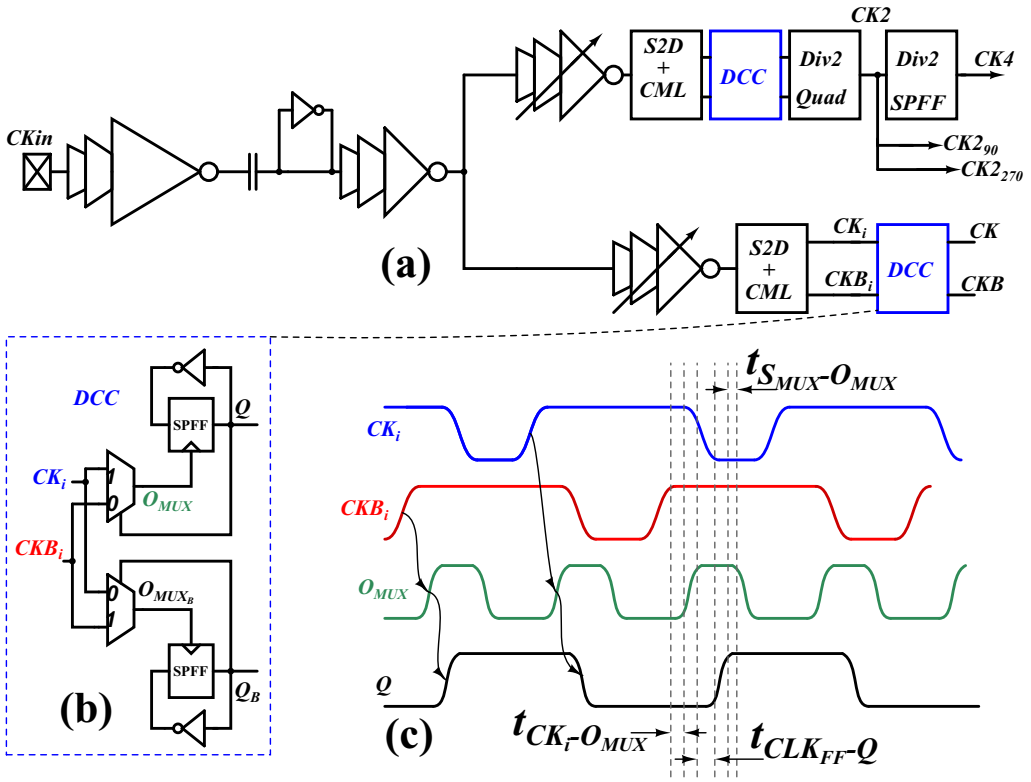


Fig. 4.23: (a) Clock distribution diagram, (b) DCC proposed circuit, and (c) DCC timing diagram.

inverted signal of the FF input and causing a transition of the divider-by-2 output configuration. The output Q (CK) transition after a delay t_{CK_i-Q} is given by:

$$t_{CK_i-Q} = t_{CK_i-O_{MUX}} + t_{CLK_{FF}-Q} \tag{4.1}$$

where $t_{CK_i-O_{MUX}}$ is the delay of the input of the mux to its output and $t_{CLK_{FF}-Q}$ is the clock to the output delay of the mux. The output of the mux O_{MUX} will transit to low once the output Q has final transition and the delay between the two events is given by $t_{S_{MUX}-O_{MUX}}$, which is the delay from the mux selector input to its output as depicted in Fig. 4.23(c). A time constraint is given by the minimum duty cycle of

the triggering signal O_{MUX}, which must not be longer than the period of the output signal (T_{CK}) by 25%:

$$t_{CK_i-Q} + t_{S_{MUX}-O_{MUX}} \leq \frac{T_{CK}}{4} \tag{4.2}$$

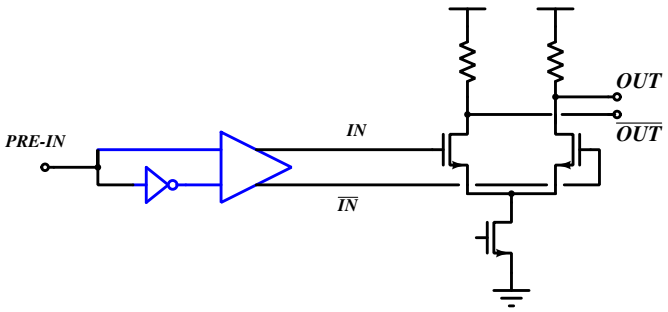


Fig. 4.24: Concept of a single-to-differential circuit.

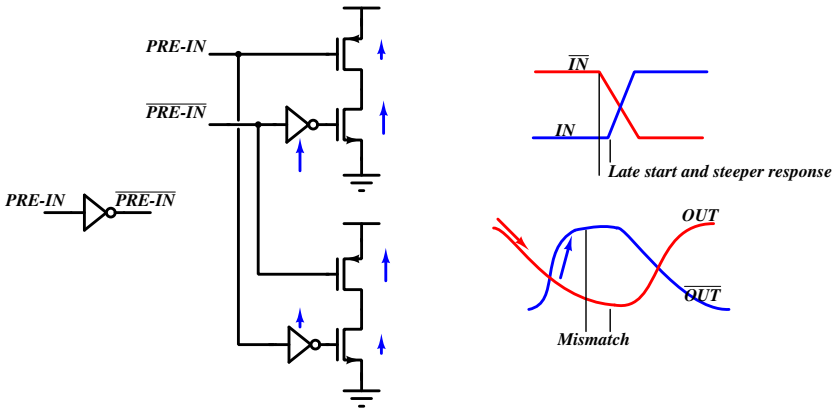


Fig. 4.25: Traditional approach of a single-to-differential circuit.

4.5.2 Single-to-Differential

The simplest S2D circuit is to use an inverter to generate the complementary phase, Fig. 4.24. However, the intrinsic inverter delay generate mismatches at the complementary output that can generate undesirable behavior to the circuits requiring

complementary phases. Furthermore, inverters have faster falling transitions due to the intrinsic characteristic of the nMOS when compared to the pMOS characteristic, as indicated in Figs. 4.25 and 4.26. In the proposed S2D depicted in Fig. 4.27, the added M_c transistor at the pull-up of the inverter tracks the complementary signal at node IND slowing the high level transition of the inverter input and allowing the pMOS transistor of the output inverter to have similar transition time with that of the nMOS device. Therefore, controlling the slewing of the falling transition at node IND allows to adjust the slewing response at the output OUT of both transitions. A programmable parallel tail to the pull-down branch of the inverter connected to node IN allows to control the falling slewing through a 4-bits digital code (slew[3:0]). Consequently, the phase matching of the differential output is improved if the S2D output is enhanced with a duty cycle correction circuit.

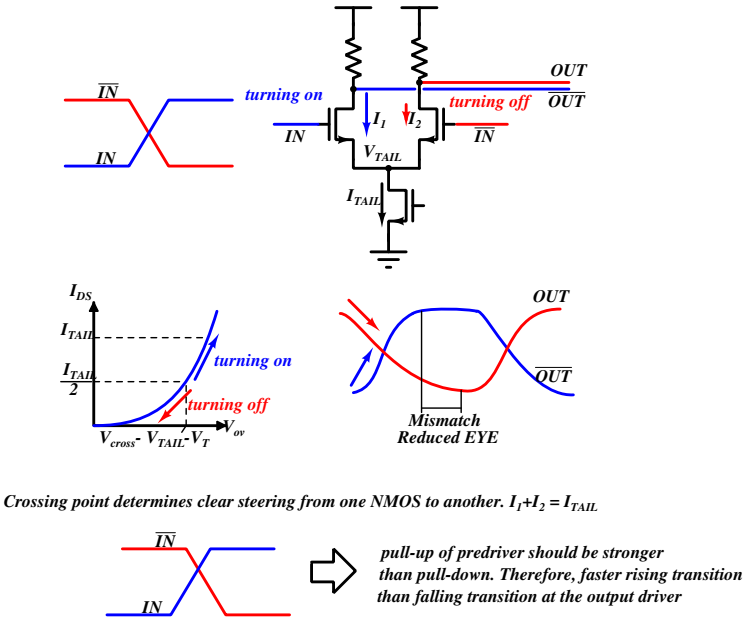


Fig. 4.26: Problems of a traditional single-to-differential circuit.

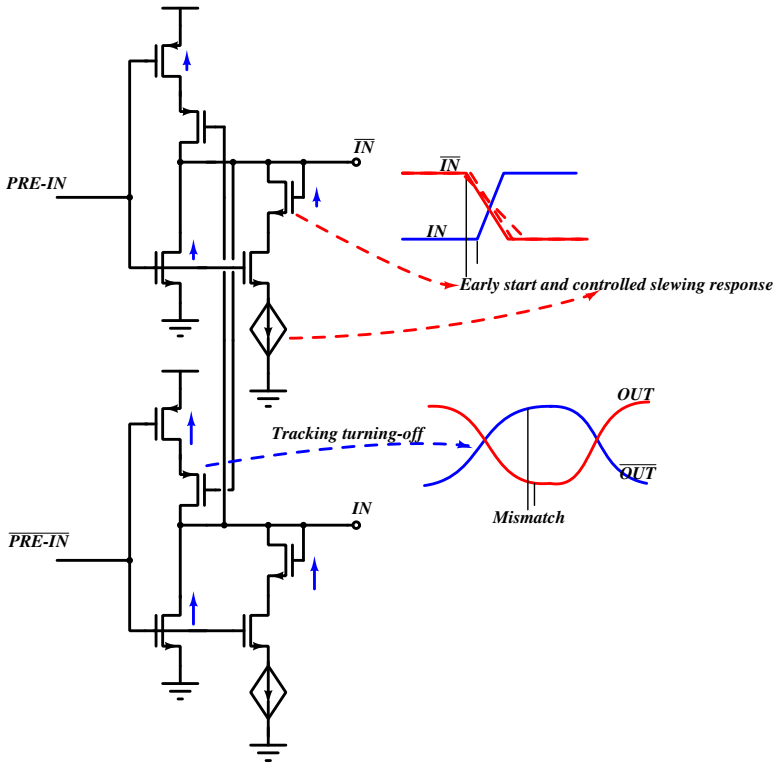


Fig. 4.27: Proposed single-to-differential circuit.

4.6 Feed-forward equalizer

At the transmitter side, the stream data is pre-distorted to achieve a flat frequency spectrum that is seen at the receiver side. The pre-distortion stage acts as a filter that attenuates low-frequency signal content and keep high-frequency signal untouched. This kind of equalization is usually referred as pre-emphasis. Pre-emphasis equalization is commonly implemented using FIR filters. In general, a FIR filter is the weighted sum of past and/or future samples subtracted from the present transmitted data. Figures 4.28 shows the concept of a 2-taps FIR. Since the output of these type of equalizer is feed-forwarded and not directly derived from the Y[n] output stream, it is called feed-forward equalizer (FFE).

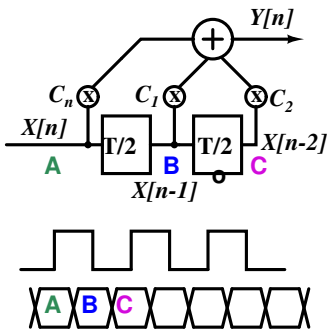


Fig. 4.28: 2-Tap FFE equalizer concept.

A conventional FFE uses passive elements with an adjustable transconductance to act as a filter. The transconductance is digitally controlled to set the specific weight values according to the channel response [32]. The differential implementation as long with the passive network introduce a power and an area overhead that is not suitable in large count I/O pins systems.

A different approach is to use active delay elements to build the FIR filter. In moderate-speed link systems, it is common to find static flip-flops as the delay elements since their operation frequency is high enough to handle moderate-speed data rate. However, in high-speed data rate systems, high-frequency boosting techniques, such as inductive peaking, are preferred for its achievable high-speed performance. Recently, CML-based latches and flip-flops have been used intensively to implement FIR filters at data rates above 10Gb/s [32]. Unfortunately, the required constant current might not be an option due to the increased power density in large I/O pins count.

True-single phase clock high-speed latches are suitable to be applied at the equalization stage. Replacing the delay elements in the FFE by true-single phase clock latches, it is possible to achieve a high-speed performance with low power consumption penalty compared to the CML-based latches. The circuit in figure 4.29 shows

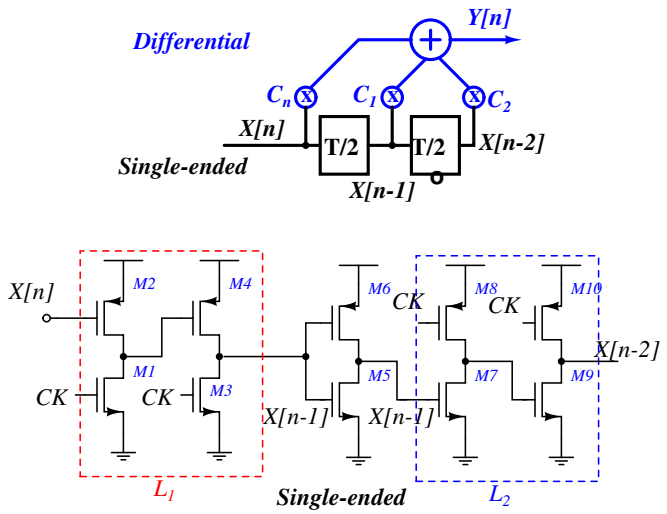


Fig. 4.29: 2-Tap FFE proposed equalizer with details on high-speed latches.

details a 2-taps FFE with details in the implementation of the latches. The latches might also be used to obtain pre-cursor taps.

4.6.1 Divider-by-2 with Quadrature Output

The 4:2 MUX works as the first serializer stage and is implemented using a standard transmission gate flip-flops (TGFF). The following MUX 2:1 uses the quarter-rate signal clock (CK2) providing a half-rate data path where the pre-emphasis taps are selected. The 2:1 MUX uses quadrature clock phases $CK2_{90}$ and $CK2_{270}$ separated by 90^o from CK2 and CK2B as indicated in Fig. 5.2(c). For the implementation of 2:1 MUX, SPFFs with a keeper in the Q_B node are used since the data input stream might have low activity. The use of SPFF saves more than 50% of power when compared to that of TGFF implementations at data activity as low as 3%, Fig. 5.2(d). The quadrature output of CK2 is obtained with the proposed domino-like based divider-by-2 shown in Fig. 4.30. A dynamic based latch with cross-coupled domino-like logic is proposed to implement the FF of divider-by-2. With proper device sizing it is possible to obtain the proper operation depicted with the timing diagram. This

implementation offers 2 and 4 folds of power savings when compared to those of a dual TSPC implementation and a CML implementation, respectively.

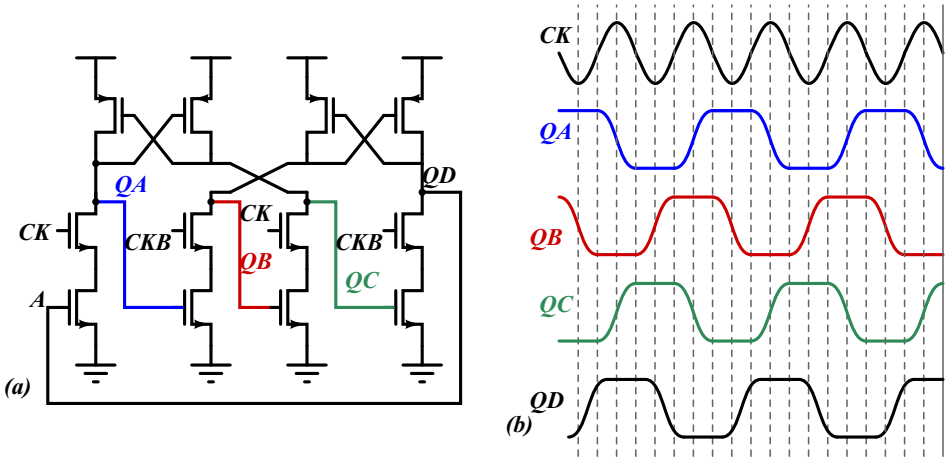


Fig. 4.30: (a) Proposed divider-by-2 with quadrature outputs and (b) its respective timing diagram.

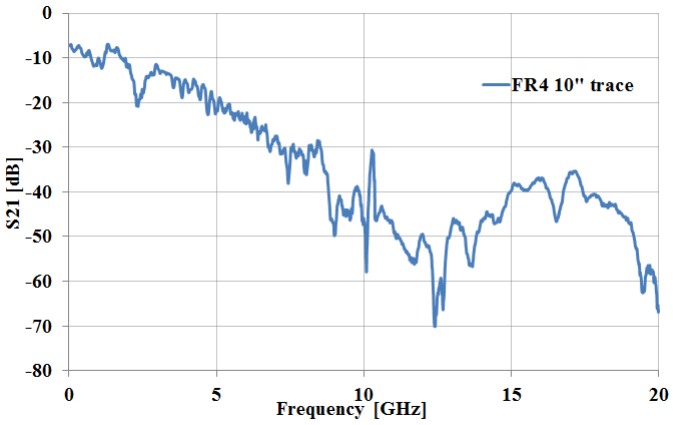


Fig. 4.31: Measured S_{21} plot of 10" FR4 trace with connectors.

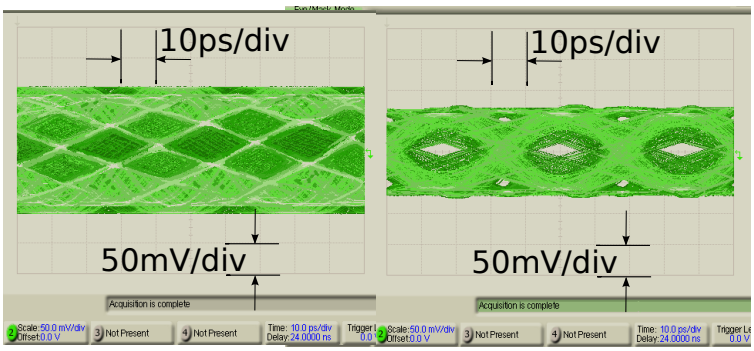


Fig. 4.32: Measured 30Gb/s differential eye with $2^{15}-1$ data pattern with average (16) enabled (a) Pre-emphasis off, (b) pre-emphasis on.

4.7 Experimental Results

In the proposed TX, all circuits use a digital standard format with a 12 track height ($2.4\mu m$) providing compact final layout and therefore a reduced load to the signal path. PRBS generators and quarter-rate circuitry were automatically placed and routed using standard digital design flow. The test-chip was wirebonded to a FR4 board with a bondwire length of about 1mm. Fig. 4.31 shows the measured S_{21} of the 10" microstrip including SMA connectors at the ends. About 50dB loss is measured at 15GHz after 10" PCB trace. Eye openings are measured using an input stream with the built-in $2^{15}-1$ PRBS sequence generator. Fig. 4.32(a) shows the measured far-end eye diagram at the connector with no pre-emphasis. Fig. 4.32(b) shows the measured eye diagram with a pre-emphasis optimized for maximum horizontal opening. Initial tap selection and weight control is executed in a PC by monitoring the error of the PRBS checker. Fig. 4.33(a) shows the measured bit error rate (BER) with the built-in PRBS checker at different data rates. For BER measurements, the PRBS generator is enabled for about 1 minute and the 15-bit error counter read it. If overflow occurs in the counter, data rate is decreased once initial tap settings are adjusted. Fig. 4.33(b) shows the measured energy efficiency at different supply VDD voltages. The driver supply (AVDD) voltage is adjusted manually and its value varies

for each VDD step. The maximum data rate recorded is based on a BER of 3.3×10^{-8}. Since the proposed TX circuits have lower numbers of cascoded transistors per branch, the TX is able to operate at lower supply voltages allowing further power savings in a system utilizing a voltage scaling strategy. Fig. 4.34 compares the breakout power of the proposed TX with a similar simulated TX including CML FFs and latches at high speed paths at different supply voltages. The total power of the proposed TX is measured and the breakout values are derived with simulations considering that the power of the driver is also measured. Both transmitters have the same driver scheme. Fig. 4.35 shows a die photograph, layout details, and the test board. Table 4.2 compares the measured performance of this work with prior works.

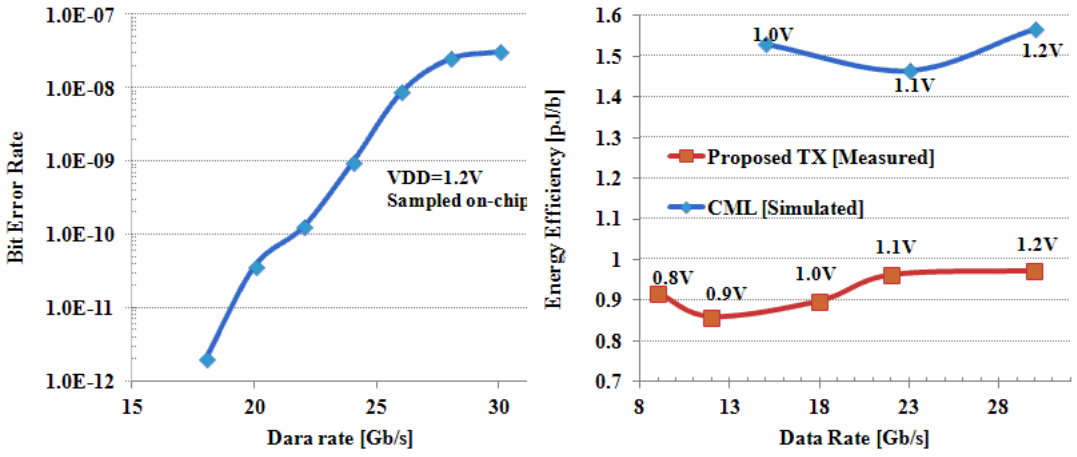


Fig. 4.33: (a) Measured BER at the output of the TX. (b) Measured energy efficiency of the TX and simulated efficiency of a TX with CML blocks.

The use of aggressive programmability in the control taps of the pre-emphasis and programmable adjusting circuits for clock skew and duty cycle corrections enable a highly efficient transmitter equalizing a channel loss of 50dB with a 30Gb/s data rate. The implemented transmitter work demonstrates that combining a single-to-differential circuit with a duty cycle correction circuit provides a significant enhancement in re-timing the final serialization stages with the presence of inevitable mismatches in complementary phases. The implementation of single-phase circuits

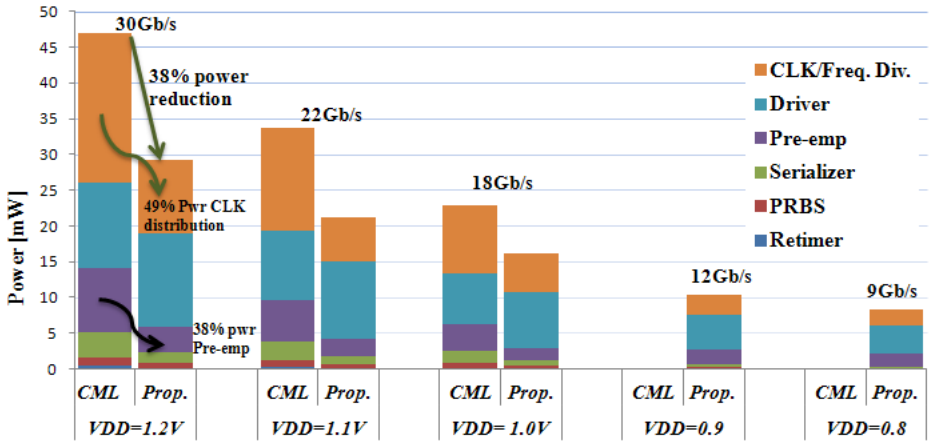


Fig. 4.34: Total power comparison of the proposed TX (measurements) with a simulated TX with selected CML blocks at different data rates.

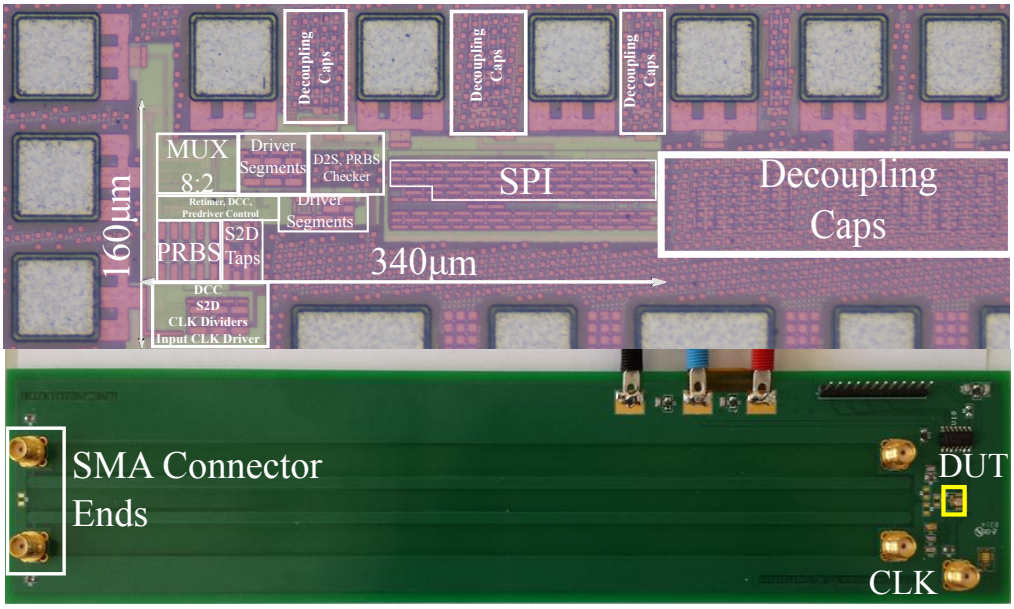


Fig. 4.35: Micrograph of TX and test board.

in most of the signal path provides a power savings of 50% in the pre-emphasis and clock distribution circuits when compared to differential implementations.

Table 4.2. Performance summary and comparison.

Ref	**Tech.**	**VDD**	**DR**	**C. Loss**	**Ener. eff.**
[30]	32nm	1.1V	28Gb/s	35dB	7.75pJ/b
[33]	65nm	1.2V	10Gb/s	13dB	1.1pJ/b
[34]	65nm	1.2V	48Gb/s	NA	1.24pJ/b
This work	65nm	1.2V	30Gb/s	50dB	0.95pJ/b

5. SINGLE-PHASE FLIP-FLOP

Timing constraints in most of the auxiliary circuits of the receiver and transmitter are limited by the delay of the FFs operating in the critical paths. For instance, in the DCC circuit, the FF must operate at half-rate which demands a high-speed FF implementation. In addition to the required high performance operation, a low-power implementation is desirable since its operation is at the highest clock rate of the TX. A true single-phase clock (TSPC) FF is attractive due to the lower power consumption compared to current mode logic (CML) implementation [35]. However, a non-negligible glitch appears at node Q_B when CK transits to high and D remains in low level, Fig. 5.1. An improved version of the FF, SPFF, is proposed and shown in Fig. 5.2 and Fig. 5.3 shows the SPFF with keeper at the Q_B node. Eliminating the pre-charge transistor alleviates the glitch issue since node B does not transit to high when CK goes to low and D remains in low level. A cross-coupled pMOS transistor is added to keep the node A high when B is low. As a result, the cross-coupled transistor works as a keeper for nodes A and B increasing the performance of the FF when input data activity is reduced. An additional advantage of the proposed SPFF is that node B keeps its logic value when CK transits and there is no signal change in the data input. This characteristic decreases the power consumption at lower data rate activity with high speed clocks.

5.1 Design Strategy

To get more insight of the proposed FF topology, a proposed a characterization setup to is presented. As a result, a SPFF that have an attractive performance for different applications is desired for comparison. Although, high speed interfaces require faster sequential elements, we have adopted the strategy of optimizing the

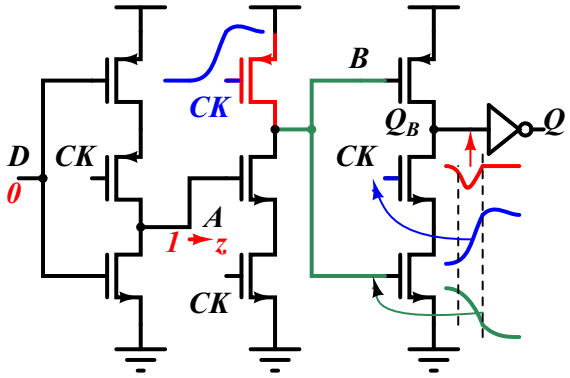


Fig. 5.1: TSPC FF and glitch illustration.

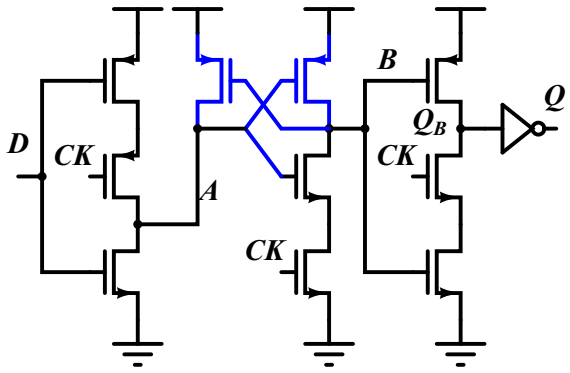


Fig. 5.2: Proposed single-phase flip-flop SPFF

design to operate in the energy efficient point to size the transistors that yield the minimum energy-delay product.

Since the slope of the clock and the data input signal might impact the power and delay measurements, it is a good practice to fix the slope to make a fair comparison with other flip-flop topologies [36, 37] . Additionally, in order to keep the standard metric for measuring digital circuit performance, we modeled the logic path between FFs as a static inverter chain with each inverter driving four copies of it itself to get a fan-out-of-four (FO4) [38]. Consequently, every gate in the delay measurement circuit sees a FO4 load yielding a slope associated to the FO4 load.

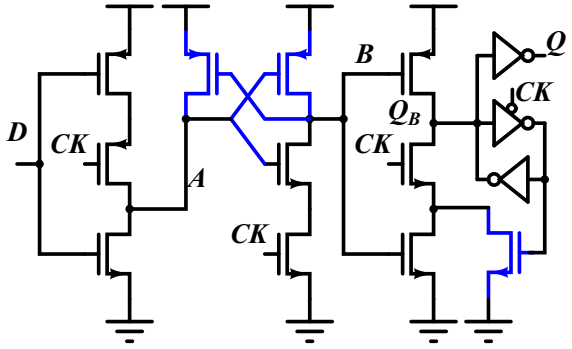


Fig. 5.3: SPFF with keeper at the output node.

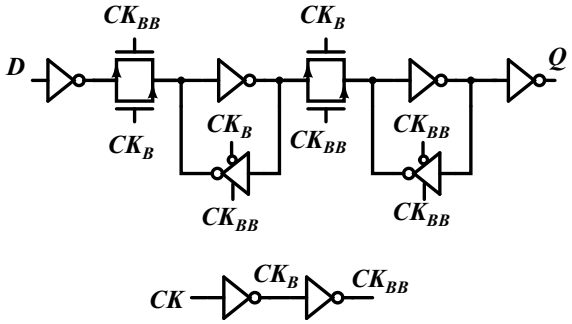


Fig. 5.4: A common transmission gate flip-flop schematic.

In order to account for the maximum operation frequency of the SPFF in a pipelined system, we have to find the optimum pipeline depth that constraints the clock frequency. Choosing the pipeline depth that produces the highest frequency is a common approach that computer architects for high-performance systems take [39–41] . Although this methodology is constrained by power, usually power is addressed by scaling the voltage or the frequency in the latest design stages such that the design fits inside the power budget. However, in a energy efficient design, is common to choose the pipeline depth that meets the minimum energy point [39]. In our approach, we chose to explore the logic depth (known also as delay of the logic circuits or the inverse of the pipelining depth) of a pipeline stage that yields the minimum

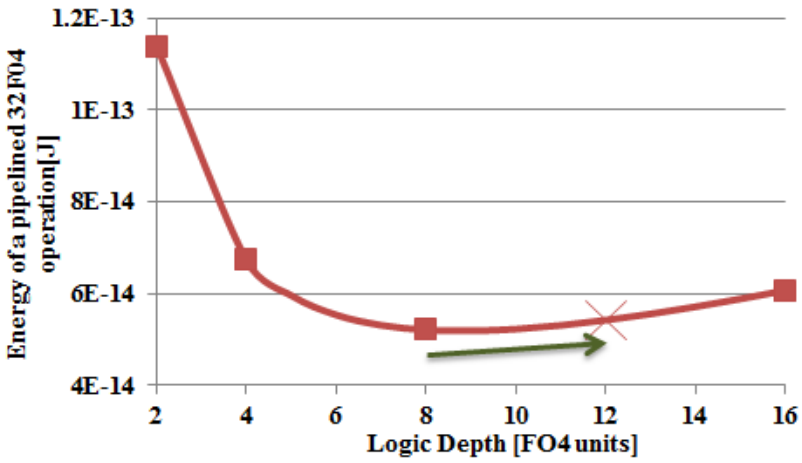


Fig. 5.5: Energy per operation of a 32FO4 logic function for different logic depths measured in FO4 delay units.

energy. The result is given in Fig. 5.5, where the x axis represents the logic depth in FO4 delay units which for the TSMC 65nm LP technology FO4 is about 25ps in the nominal case. We have assumed the following to calculate the logic depth: (1) a fixed load of FO3 that each gate sees, (2) a voltage supply of 1.2V, (3) a switching activity of 25% of the data input stream, (4) and the SPFF sized for minimum E-D product. A logic depth of 6FO4 results in the minimum energy point. This result follows the trend that a low logic depth reduces the idle time of the gates between the FFs and therefore lowers the leakage energy per cycle of the gates. Whereas with lower logic depth, the number of FFs increases and the clocking overhead and higher frequency increases the dynamic energy [40].

In regard to process variations, a common approach is to consider longer logic paths to reduce the effect of gate variations on total path delay by exploiting the delay averaging characteristic. For instance, if we choose 8FO4 as the logic depth will give us a clock frequency of 5GHz. Operating at 5GHz, a 100 points Monte Carlo analysis showed over 25% variations in the total delay path of a pipeline stage. Assuming that 25% of delay variation might not satisfy the multiple worst-case delay paths in a full system-on-chip architecture, different Montecarlo analysis were run for

operation frequencies below 5GHz until we found a total path delay variation within 10%. At 3.3GHz operation frequency, the logic depth is about 12 FO4 and the total delay path variations per stage stays within 10%. With a logic depth of 12FO4, a simulation setup of 3 pipeline stages with a 3.3GHz clock was used to get simulation results of FF power consumption for different switching activity.

5.2 Characterization circuits

In order to set a performance framework, it is necessary to define the strategy to obtain the different delay and energy measurements. The clock-to-output delay (t_{C-Q}) is the propagation time between the triggering clock edge and the new output transition given that the data signal has been set earlier than the clock edge. As the data signal edge gets closer to the triggering clock edge (t_{D-C} decreases), t_{C-Q} starts increasing and it might rise exponentially. If the data edge is pushed even closer to the clock edge, the data transferring to the output might fail. Therefore, a enough setting time of the data is required before the triggering clock edge. The minimum setting time ($t_{D-C_{min}}$) that allows to reliable transfer the data to the output is defined as setup time (t_{setup}). Since the optimization interest is to reduce the propagation delay of the data to the output (t_{D-Q}) terminals of the FF, it is important to be able to reduce the data-to-clock and clock-to-output delays since $t_{D-Q} = t_{D-C} + t_{C-Q}$. The minimum t_{D-C} is the setup time and is commonly defined where t_{D-Q} minimum occurs. It means that $\frac{dt_{D-Q}}{dt_{D-CLK}} = 0$ which occurs when $\frac{dt_{CLK-Q}}{dt_{D-CLK}} = -1$.

After the triggering clock edge occurs, the data signal must be held constant to allow the output to transition. The minimum time that the data signal must be constant is called the hold time (t_{hold}). The hold time becomes critical when it equals to t_{C-Q} plus the time it takes of data to arrive at the next FF stage which in pipeline paths is due to the logic between the FFs, and its propagation delay is defined as t_{logic}. Assuming that t_{logic} is zero and there is not clock skew between stages, a system will operate correctly if we assume that the t_{C-Q} obtained at the setup measurement

is the same time obtained for the hold time. Therefore, is common to define the hold time as the time difference between data arrival and the triggering clock edge that has the same t_{C-Q} delay of the t_{C-Q} calculated for the setup time.

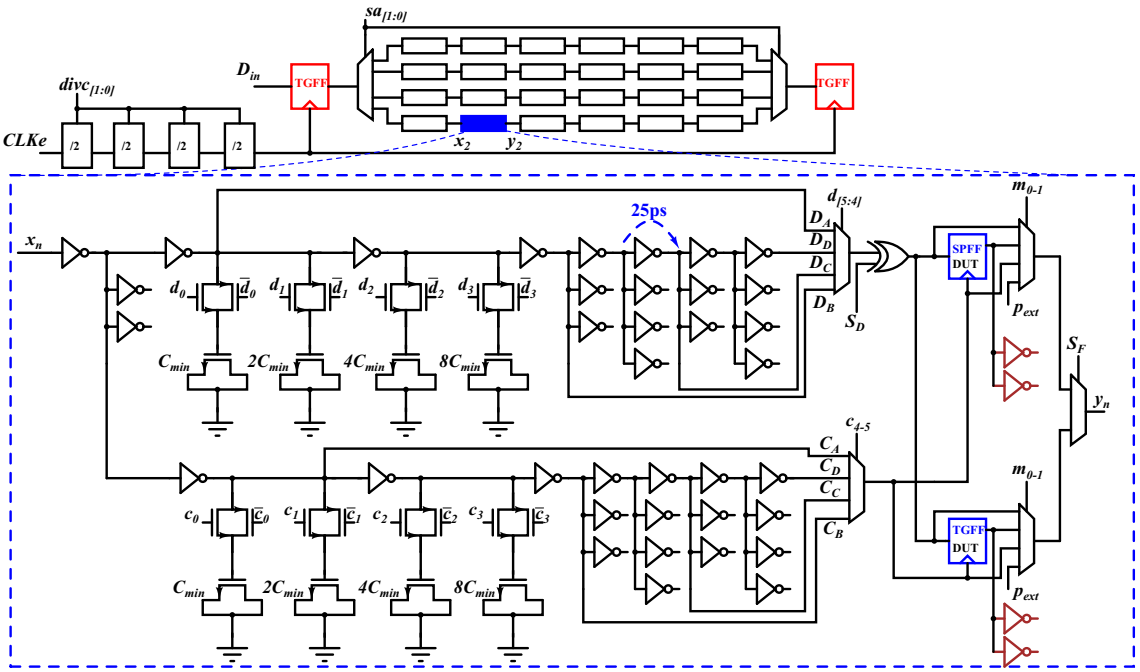


Fig. 5.6: Implemented test circuit to measure delay characteristics of the flip-flops TGFF and SPFF.

5.2.1 Delay testing setup

In order to measure the delay characteristics of the proposed SPFF, the testing setup shown in Fig. 5.6 was implemented. An array of six SPFFs per row and four SPFFs per column were implemented. The array scheme allows to measure the characteristics by averaging. Furthermore, the proposed scheme allows to extract performance characteristics with requiring additional on-chip circuitry that might impact the accuracy of the measurements. For instance, compared to [42] which depends on the accuracy of a ring oscillator, the proposed structure uses an external

accurate clock source. In order to extract the timing characteristics, the following procedure is required:

1. Set $m_{[1:0]}$ of each stage to select the output of flip-flop under test (FFUT).

2. Set $d_{[5:4]}$ of each stage to select D_A. Set $C_{[5:4]}$ to select D_D. This result in an initial maximum t_{D-C} per stage.

3. Flush the pipelined array by setting a logic value 0 at the input.

4. Set D_{in} with a serial input pattern 0011001100110011 synchronized to a clock of 100MHz. Set CLK_e to 1600MHz and enable the four dividers-by-2.

5. Store the output stream D_{out} and check that the pattern 0011001100110011X appears.

6. Increase the frequency of the signal clock that the extreme TGFFs see (CLK_A) by bypassing one-by-one of the dividers-by-2. In every increasing step, check that the pattern 0011001100110011X appears. If the output stream D_{out} fails to capture the pattern correctly, save the divider configuration. At this point, the period (t_{CLK_A}) of signal CLK_A is smaller than the total propagation delay of a row of the array (t_{darray}) unabling to capture properly the data at D_{out}. Let us called the period t_{CLK_A} of failing to capture D_{out} as t_{CLK_AF}. The propagation delay of a row of the array can be calculated approximately by:

$$t_{darray} = 6 * (8 * FO4 + 3 * t_{MUX} + t_{D-Q} + t_{XOR}) + 2 * t_{MUX-A} \tag{5.1}$$

where FO4 is approximately 25ps, t_{MUX} is the propagation delay of the 4-input mux and is approximately 80ps, t_{XOR} is the propagation delay of the XOR gate used to select the sign of the input slope to the FFUT, and is approximately 80ps, t_{MUX-A} is the delay inserted by the mux selecting the array rows with an approximated value of 100ps, and t_{D-Q} is the propagation delay from the data input to the output of the FFUT. The approximate total delay is about 3.8ns.

7. Select the t_{CLK_A} equals to $\frac{t_{CLK_AF}}{2}$.

8. Set $d_{[5:4]}$ such that the t_{D-C} of FFUT is decreased until the input pattern is not captured correctly. If the pattern still correct after having changed the selecting taps of the data path of the FFUT, modify the slope gradually by changing $d_{[3:0]}$ and having $d_{[5:4]}$ selected to D_A.

9. Similar to the previous step, it is possible to modify t_{D-C} by bringing the clock edge close to the data signal edge by setting $c_{[5:4]}$.

10. Setting t_{D-C} of FFUT by either one or both of the two previous steps, an additional accuracy is obtained by increasing the external clock (CLK_e) frequency progressively until D_{out} stream fails to capture the pattern correctly. Given the settings of the dividers-by-2 and the input clock frequency, t_{D-C} is calculated as:

$$t_{D-C} = t_{C-fail} - t_{D-fail} \tag{5.2}$$

where t_{C-fail} is the arrival time of the clock edge and t_{D-fail} is the arrival time of the data edge when D_{out} starts failing to capture the input pattern.

11. t_{D-fail} is obtained by setting $m_{[1:0]}$ of each stage to select the input of the FFUT. Then, increasing the external clock frequency progressively until D_{out} stream fails to capture the pattern correctly. t_{D-fail} is calculated as:

$$t_{D-fail} = \frac{t_{CLK_AF}}{6} \tag{5.3}$$

12. t_{C-fail} is obtained by setting $m_{[1:0]}$ of each stage to select the clock terminal of the FFUT. Then, increasing the external clock frequency progressively until D_{out} stream fails to capture the pattern correctly. t_{C-fail} is calculated as:

$$t_{C-fail} = \frac{t_{CLK_AF}}{6} \tag{5.4}$$

Similar to t_{D-fail} calculation, t_{C-fail} contains the propagation delays of the muxers and the XOR gate and therefore they canceled out at the calculation of t_{D-C}.

13. t_{C-Q} is calculated as:

$$t_{C-Q} = t_{Q-fail} - t_{C-fail} \tag{5.5}$$

where t_{Q-fail} is obtained by setting $m_{[1:0]}$ of each stage to select the output of the FFUT. Then, increasing the external clock frequency progressively until D_{out} stream fails to capture the pattern correctly. t_{Q-fail} is calculated as:

$$t_{Q-fail} = \frac{t_{CLK_AF}}{6} \tag{5.6}$$

14. Repeat all previous steps after selecting a different row of the array by setting $sa_{[1:0]}$. The final t_{D-C} value is obtained after averaging the results of the four rows.

The previous procedure is also applied to calculate the hold time t_{hold}, but instead of the clock edge arriving first than the data edge, now the data edge is set to arrive first by properly setting $d_{[5:0]}$ and $c_{[5:0]}$.

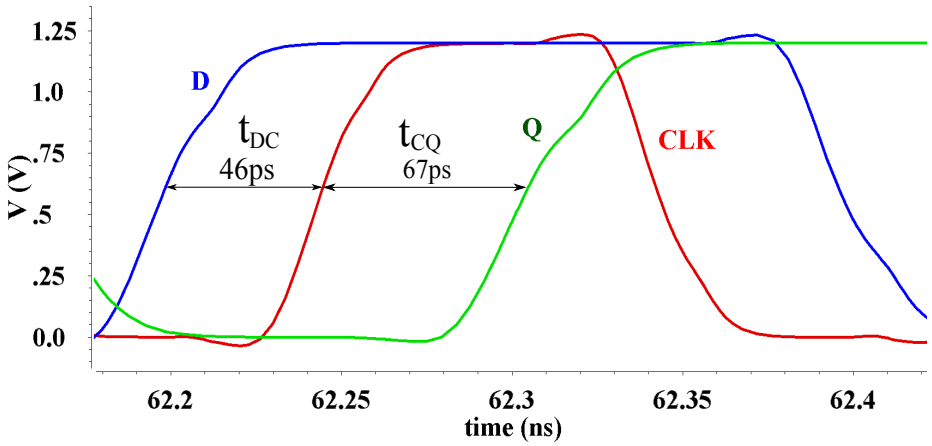


Fig. 5.7: Waveforms operation of the SPFF.

By following the measuring policies for t_{setup} and t_{C-Q} described at the beginning of the section, the simulated t_{setup} result is 46ps and the t_{C-Q} is 67ps. Figure 5.7 shows the simulated waveforms at the terminals of the SPFF of the implemented delay testing circuit in a typical corner with a t_{setup} of 46ps. Figures 5.8 and 5.9

shows the calculated setup and hold time after a simulated Montecarlo analysis with 100 points. As a result, the t_{setup} obtained is 45.7ps with a σ of 11ps, which represents an upper bound for 3σ of 78.7ps. The t_{C-Q} obtained is about 69ps with a σ of 10.2ps, which represents an upper bound for 3σ of about 100ps.

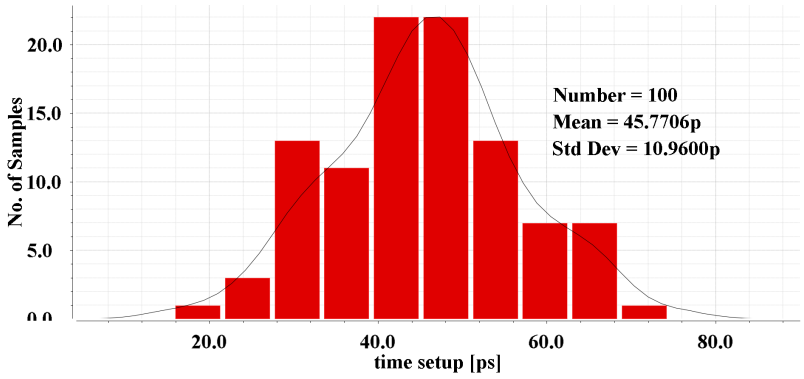


Fig. 5.8: Setup time calculated with a 100 points simulated Montecarlo analysis.

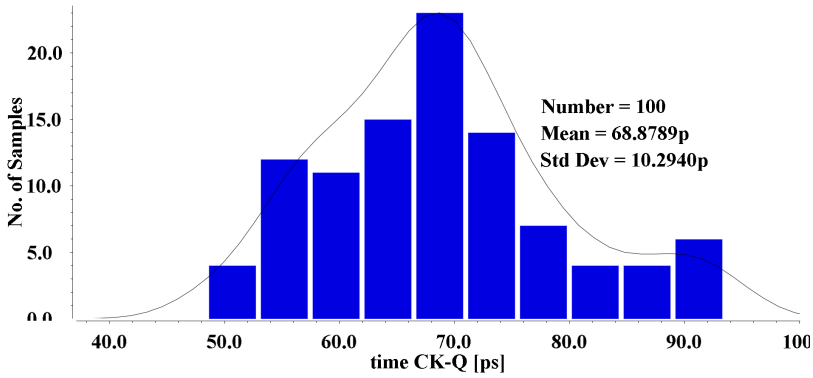


Fig. 5.9: Clock-to-output delay calculated with a 100 points simulated Montecarlo analysis.

5.2.2 Energy testing setup

The implemented test circuit to measure the energy performance of the TGFF and SPFF is shown inf Fig. 5.10. A conventional ring counter with TGFF elements is used to provide the input stimulus to four FFUT. The 8-bits activity control word sets the initial condition of the ring counter and therefore the activity seen by the flops under test. An advantage of this procedure is that the measured energy is well averaged since each flop sees different input pattern at given time but with the same activity at the total count. Slopes and load are adjusted to be FO4 and the supply of the activity control circuit is connected to the supply of the digital control interface. The supply of the FFUT, as well as the supply for the inverters and the level shifter driving the clock and data terminals respectively, are connected to an isolated VDD terminal called VDT in order to measure the supplied current directly.

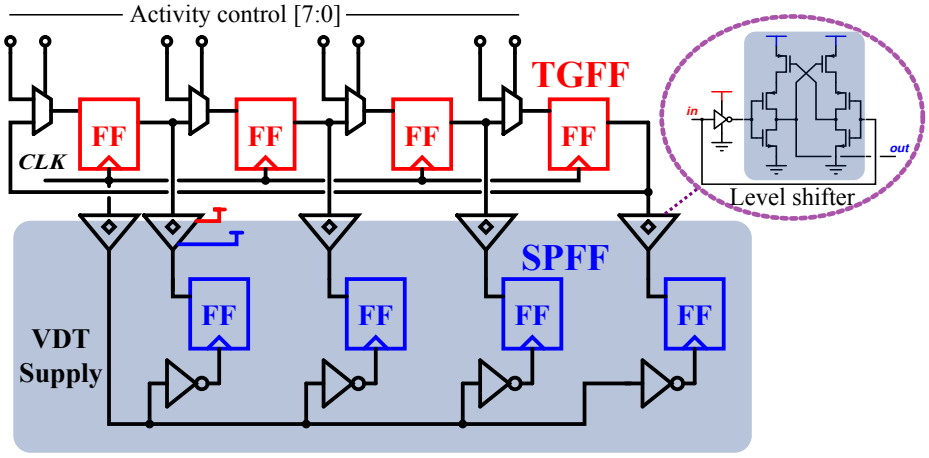


Fig. 5.10: Implemented test circuit to measure energy performance.

5.3 Experimental Results

The procedure to calculate the timing characteristics of the implemented TGFF and SPFF is applied using a script written in python. Data is send/receive through

the USB port of the PC and a USB-to-SPI cable connected to the SPI port of the chip. Lab data is gathered and processed to generate the plot shown in Fig. 5.11. When the data at the output fails to capture the input, the t_{D-Q} and t_{C-Q} values were kept the same as the latest value when the data was captured correctly. The setup time obtained for the implemented TGFF is 63ps and 133ps for t_{C-Q}. For the SPFF time characterization, the t_{D-Q} does not show a convex plot and consequently the adopted definition to measure setup time does not apply directly. Given the simulation results of 47ps for t_{setup}, the slope of the t_{C-Q} curve for t_{D-C} at 47ps is different than the slope of t_{C-Q} when the minimum of t_{D-Q} occurs for the TGFF measured case. We use the same slope of t_{C-Q} obtained for TGFF at given t_{D-C} to keep a similar timing margin. The measured setup time for the SPFF is about 50ps and t_{C-Q} about 65ps.

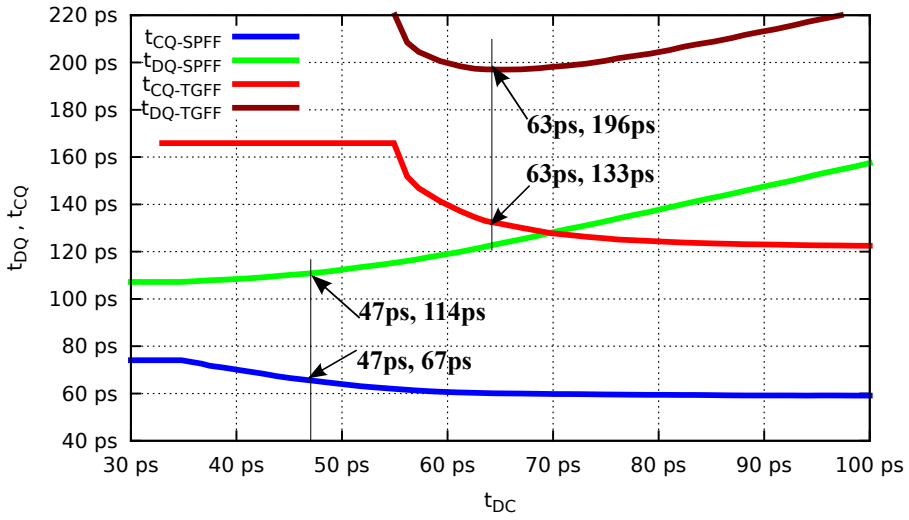


Fig. 5.11: Measured t_{setup} (t_{D-C}) and t_{C-Q} of the TGFF and SPFF.

Figure 5.12 shows the measured power consumption at different activity rates when the provided clock frequency is 3.3GHz. Compared to the TGFF, a 45% power savings is achieved at 25% data activity. Considering that in a worst case scenario for interface applications or pipelined operations data activity might not get as high

as 25%, this result shows a strong feasibility for replacing conventional clock storage elements in restricted power-budget applications.

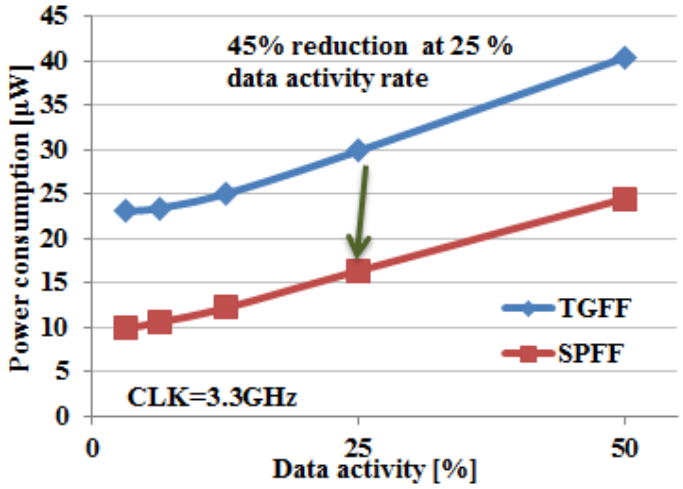


Fig. 5.12: Measured power comparison of the proposed SPFF and the TGFF for different activity rate with clock rate of 3.3GHz.

Figure 5.13 shows the details of the die with the implemented testing structures in a TSMC 65nm low power logic CMOS technology. Table 5.1 compare the performance of the SPFF and TGFF with recent reported work. Although the comparison is unfair since the delay and energy characteristics depend on technology, load, slope, activity, clock frequency and other measurement conditions, the results of other work allow to have an insight of the performance of the proposed SPFF. SPFF has the lower number of transistors and timing and energy performance are competitive numbers among similar measurement conditions. When compared to the TGFF the SPFF outperforms with a 45% energy savings at 25% activity rate, 13ps faster for setup time and two folds faster for the clock to output propagation delay.

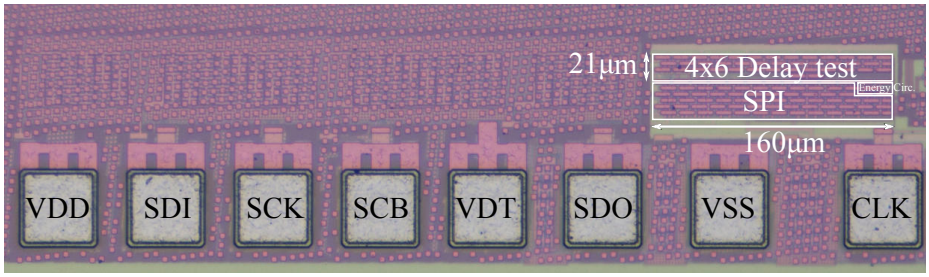


Fig. 5.13: Micrograph details of delay and energy test structures.

Table 5.1. Performance summary and comparison.

Ref	**Tech.**	**VDD**	**#Trans.**	t_{setup}	t_{C-Q}	t_{hold}	**Energy.**
[43]	40nm	1.1V	22	197ps	67ps	73ps	2fJ [a]
[44]	65nm GP	1.0V	25 [b]	29ps	49ps	17.8ps	42fJ
[45]	40nm	0.9V	48	5ps	37ps	NA	5fJ [c]
[46]	40nm	1.1V	21	105ps	176ps	69ps	3.1fJ [d]
TGFF	65nm LP	1.2V	24	63ps	133ps	18ps	9fJ
SPFF	65nm LP	1.2V	18	50ps	65ps	44ps	5fJ

[a] Measured at 10% activity rate.

[b] A shareable pulse generator is required.

[c] Measured at 50% activity and 0.81V supply.

[d] Approximated value based on 34% the power consumption of TGFF at 10% activity rate.

6. WIRELINE RECEIVER CIRCUIT TECHNIQUES

In lossy channels, the signal amplitude is attenuated as a function of frequency. Therefore, transmitting a high-speed pulse is deformed and its transition tail may be misinterpreted by the receiver with a symbol that was previously transmitted. For instance, after a long series of pulses representing logic one values, a significant amount of voltage might remain on the interconnect over several unit intervals (UI).

The previous problem is resolved employing equalization systems. Passive equalizer and continuous time linear equalizer (CTLE) networks are among the most popular. The common approach is to invert the frequency response of the channel such that the overall response is flat over the band of interest. As a result, a high-frequency boosting is required as long with a low frequency attenuation. Unfortunately, this technique introduces an additional insertion loss at the link and amplification might be required. Boosting the high-frequency content of the signal increases crosstalk and high-frequency noise. In addition, passive and CTLE equalization only compensates frequency-dependent loss and not intersymbol interference due to reflections.

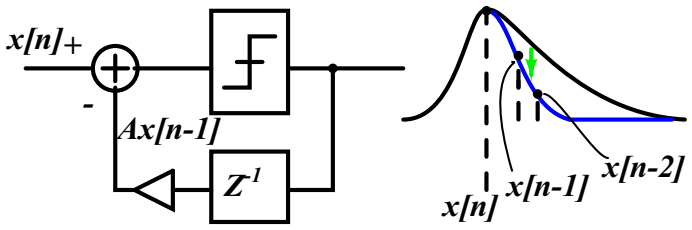


Fig. 6.1: DFE concept.

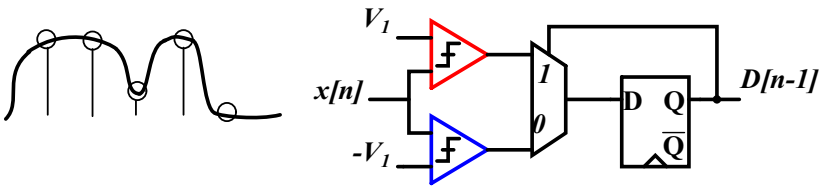


Fig. 6.2: Partial response DFE.

6.1 Decision Feedback Equalizer

In contrast to passive equalizer and CTLE, a decision-feedback equalizer (DFE) compensates based in the data content. A digital output is compared with a specific threshold to generate a digital value. This value is delayed and scaled so it can be subtracted from the input signal. Considering the lossy nature of channels, low frequency content is also attenuated which demands the need for additional gain of the low frequency. In lossy channels, it is common to have a variable gain amplifier (VGA) and a CTLE before the DFE to have the required swing at the input of the DFE. Figure 6.1 presents a 1-tap DFE block diagram serving to illustrate the concept of compensating the current input value with a fraction of the previous bit. The delay is synchronized to set its output to the previous bit, which is scaled by a factor such that post-cursor effect is canceled from the current transmitted bit.

A challenge in the design of the DFE is ensuring that the feedback signal settle within one bit period. At high data rates, this timing constraint becomes more difficult to fulfill. A technique known as loop unrolling, also called partial response or speculation, is commonly used to relax this timing constraint [47–49], fig. 6.2. Avoiding the large capacitance of the summing node and keeping the decision within the digital domain, the delay of the feedback is reduced. Previous decisions are made by unrolling the possible combinations of the previous bits. In 1-tap DFE, two combinations are unrolled by adding +V1 and -V1 DC offsets to the data input and

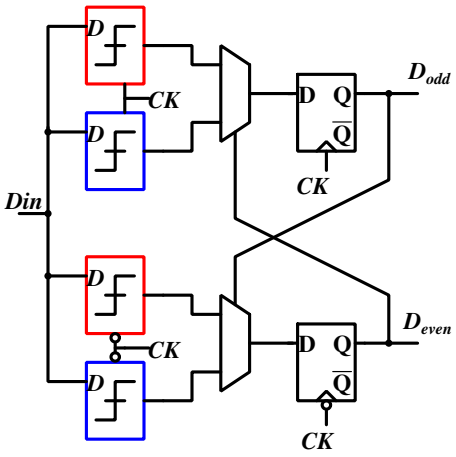


Fig. 6.3: Half rate partial response DFE.

the results are sliced to binary values. Once the previous bit is known, a MUX selects the corresponding output.

Two branches of DFEs, one running at the positive half-cycle of the clock and the other one at the negative half-cycle of the clock, are called half-rate DFE systems [49], fig. 6.3. The advantage of running at half the clock speed saves power in clock generation, but at the overall increasing in area and power consumption. Another advantage of the half-rate is the inherent DEMUX 2:1 operation at the output, consequently eliminating the need for a high-speed DEMUX. However, the fan-in of the input node increases the capacitance seen by the previous stage, requiring additional power consumption. Furthermore, the additional muxers and slicers added to the data path decreases the timing margins of the unrolled tap resulting in a timing constraint of:

$$t_{D-Q} + t_{MUX} < UI \tag{6.1}$$

where t_{D-Q} is the data to output delay of the flip-flop, t_{MUX} the propagation delay of the mux and UI the unit interval of a symbol. The increasing number of slicers seen by the data input node will load considerably the previous block restricting the maximum achievable bandwidth.

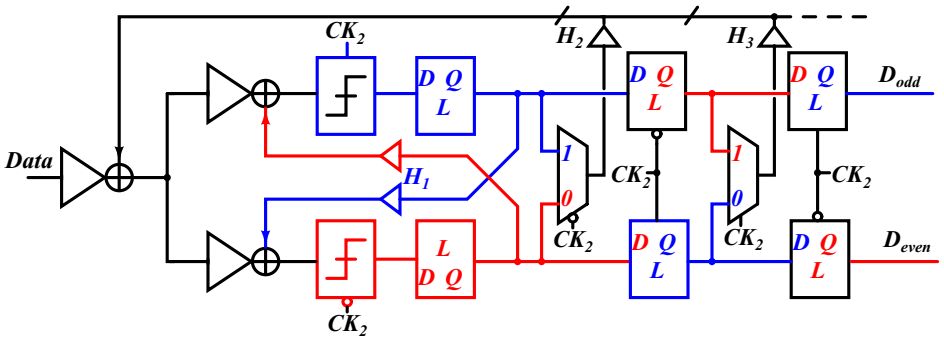


Fig. 6.4: Half rate direct closed-loop DFE.

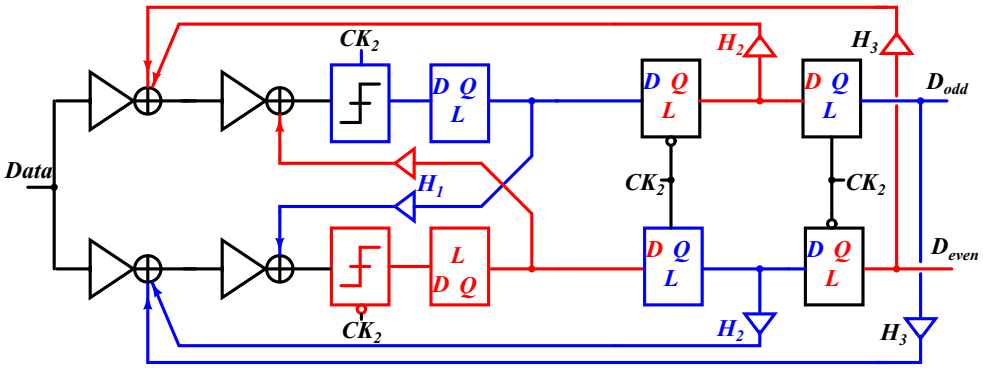


Fig. 6.5: Half rate direct closed-loop DFE without muxers to relax timing.

A conventional half-rate closed-loop, with a direct feedback as shown in Fig. 6.6 might overcome the issues of an unrolled DFE but with a cost. The burden of requiring a summer circuit for the first tap might limit the timing constraint even worse if the summer node is not isolated from considerable loads. As shown in Fig. 6.6, recent work isolate the first tap from later taps to improve the timing performance of the first tap path and achieve higher data rate. However, the additional amplifier and

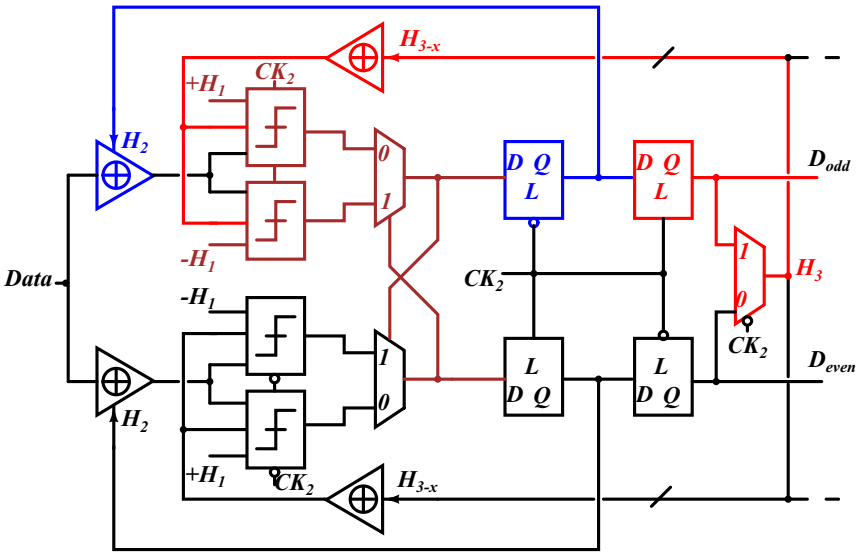


Fig. 6.6: Half rate 1-tap partial response with 1-tap closed-loop DFE.

summer circuits impact negatively the power budget. Interfaces with tighten timing specifications might require a faster second tap and therefore further optimizations are applied. Fig. 6.5 improves the timing margins of the later taps but eliminating the muxers at the cost of a required summer per half-rate branch and bigger input capacitance load of the DFE.

Recent published work [50, 51], have made use a combination of a first-tap unroll and the direct closed-loop for the second tap. However, the still larger number of slicers and summers impact power and area. In order to balance timing and power consumption performance towards energy efficient links, we proposed the circuit shown in Fig. 6.7. Although the number of flops correlated to the input node are reduced, the architecture seems to suffer of a tighter timing constraint since two muxers are in the feedback of the first tap. Interestingly, if the pair of flops and muxer are merged together, the timing constraint should be relaxed. Since both of

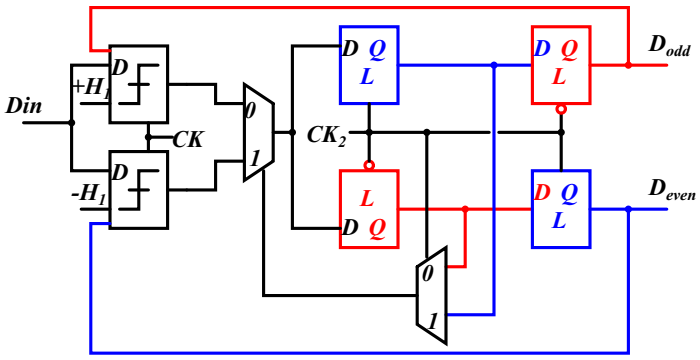


Fig. 6.7: Proposed half rate partial response DFE.

the half-cycle feedback signals are multiplexed to select the unrolled decision, we get additional power and area savings due to the reuse of summers and slicers.

6.2 Clock and data recovery

Three common approaches for clock recovery architectures are shown in figure 6.8. The one loop approach adjusts the VCO output frequency and phase to the center of the UI of the incoming data. However, the high accumulated jitter in the incoming data might be tracked if the loop-bandwidth of the PLL/DLL is high. Considering that a high loop-bandwidth is required to decrease the phase noise of the loop, it is difficult to avoid the jitter tracking behavior. In order to break the phase noise and jitter trade-off in the one loop based architecture, it is commonly preferred to have two loops CDR scheme [47]. One loop is used to adjust to the frequency of the incoming data with low phase noise, and the other loop to setup the sampling clock phase to the center of the eye data.

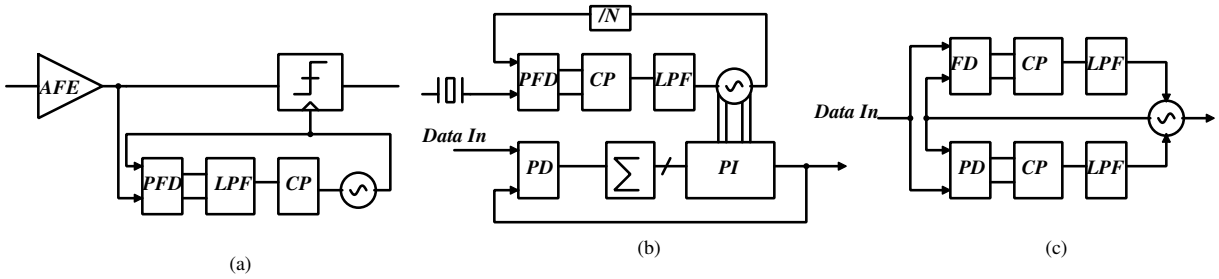


Fig. 6.8: CDR architectures: (a) Conventional one loop VCO-based, (b) Dual Loop with PI, (c) Dual loop reference-less.

Dual loop CDR architecture avoid the use of two oscillators by using a phase interpolator (PI). A phase-interpolator uses a finite state machine (FSM) to select the optimal phase from a multiple phase generator. However, generation of multiple phases introduces a prohibited overhead in power consumption at considerable high speed data rates. Moreover, two loops scheme has the disadvantage of having two different control paths that can suffer from concurrent corrections. Besides the additional complexity, recent publications [47, 52] report an intensified utilization of the dual loop architecture with PI at the expense of larger power consumption. Among the dual loop schemes, a reference-less architecture is also reported [52]. CDR without external clock reduces cost at the detriment of increase complexity and power consumption. Recently, Jung and Razavi [53], reported a 25Gb/s CDR with twenty-fold reduction in the power consumption with respect to related art. The authors employ charge-steering logic latches instead of CML based latches in a single loop CDR to achieve the reported performance. A class-C VCO with low phase-noise and a 6MHz loop bandwidth is implemented to achieve a relative jitter tolerance.

In this work, we propose the use of the implication cells to implement the required latches in the one loop CDR architecture. Since proposed latches operate with a single-phase clock, the estimated power consumption is lesser than the charge-steering based latches reported in [53]. In addition, the exploits of a merged DFE and phase detector scheme provide further power saving. Figure 6.9 presents the proposed phase

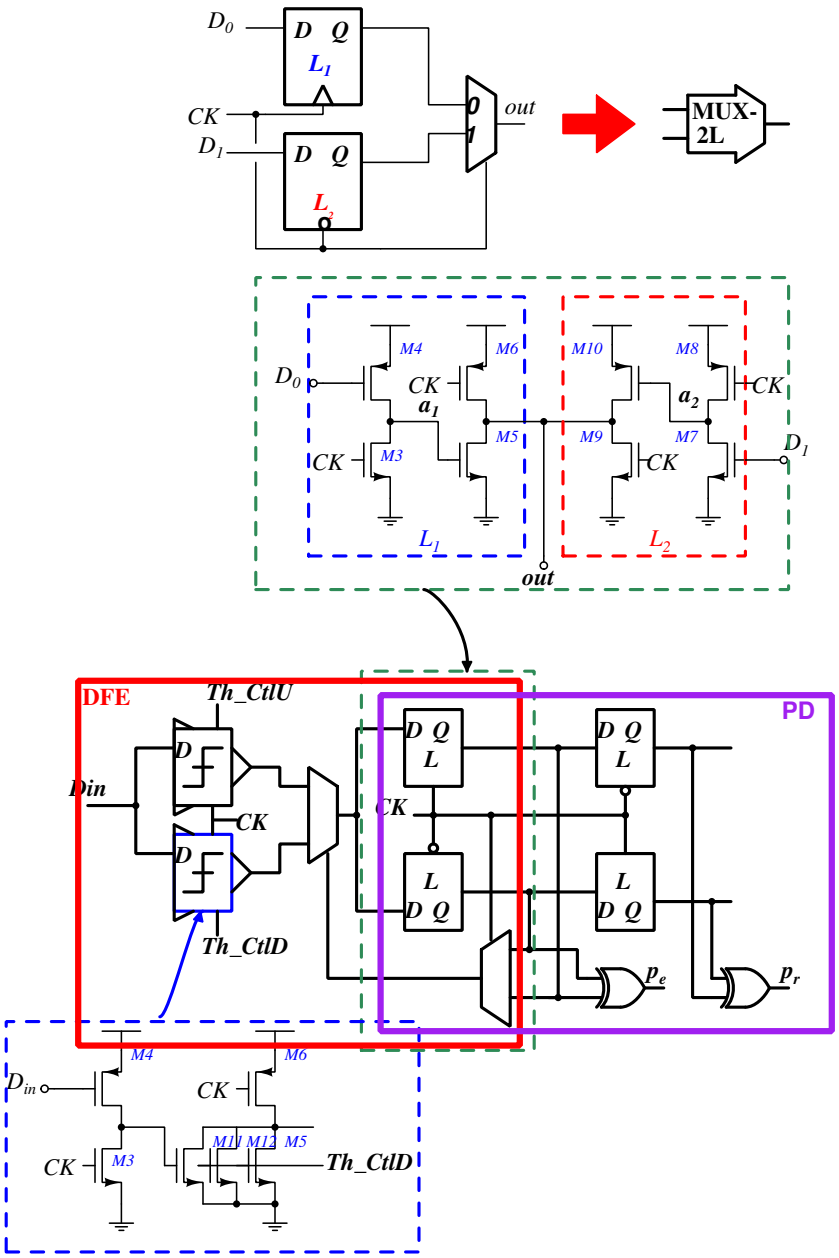


Fig. 6.9: Proposed CDR phase detector reusing part of the DFE.

detector that reuses part of the DFE as long with the details of the latches. The final proposed DFE and CDR is presented in figure 6.10. Further analysis is required to achieve the jitter tolerance at high-speed data rates. Layout implementation and

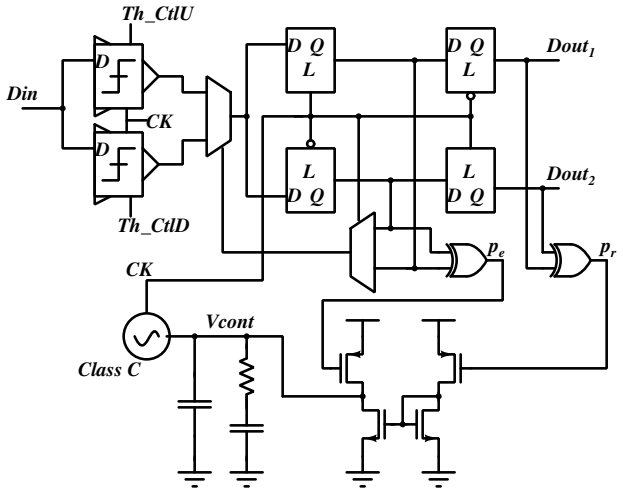


Fig. 6.10: Proposed DFE and CDR.

RC extracted simulations are also required to demonstrate the feasibility at high frequency operation.

6.3 Experimental Results

The measurement setup used to characterize the DFE and CDR is shown in Fig. 6.11. The on-chip PRBS generator with a $2^{15} - 1$ sequence is applied to the input of a configurable 3-stages gm-C low-pass filter. The low-pass filter is set to emulate the combination of a 3-taps transmitter pre-emphasis, a channel response with a 35dB loss at 12.5GHz, and a linear equalizer with a 11dB equalization.

A chip photo of the DFE and CDR is shown in Fig. 6.12. Same CDR and DFE was applied to an optical receiver pictured in 6.13. The testing structures were implemented in TSMC 65nm general purpose logic CMOS technology. A GSGSG probe was used to access the odd and even data output signals at half rate. A GSG probe was used to apply the input clock to the PRBS generators. A DC probe head with multiple contacts were used to supply power to the digital control circuits, the

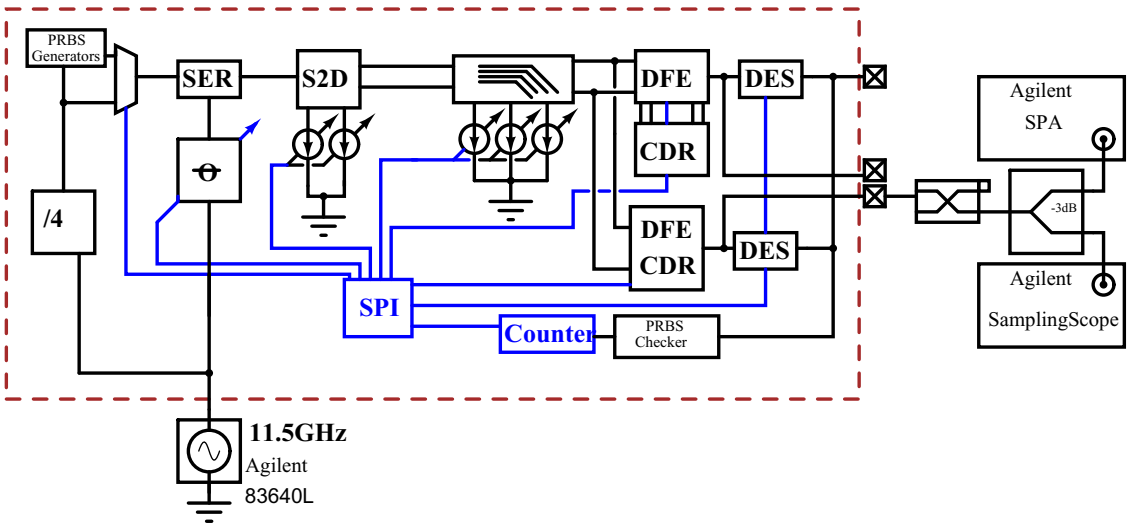


Fig. 6.11: Measurement setup for DFE and CDR chip.

filter, the VCO, and the SPI control bits. Single DC parametric probes were used to provide access to the reference bias resistors and the VCO loop filter.

The measured recovered half-rate data is shown in Fig. 6.14. Although the measured jitter is 0.31UI, the result indicates that the DFE taps are improving the timing margin and opening of the eye. A 7.8mW power consumption was measured from a 1.2V supply.

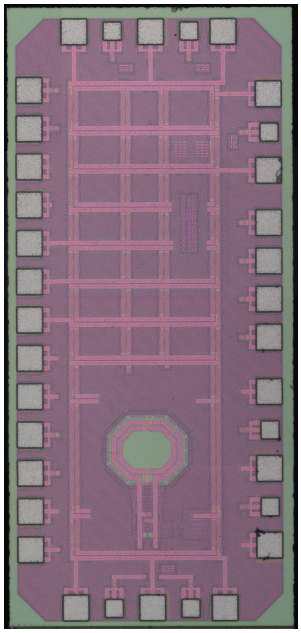


Fig. 6.12: Microphotograph of DFE and CDR chip.

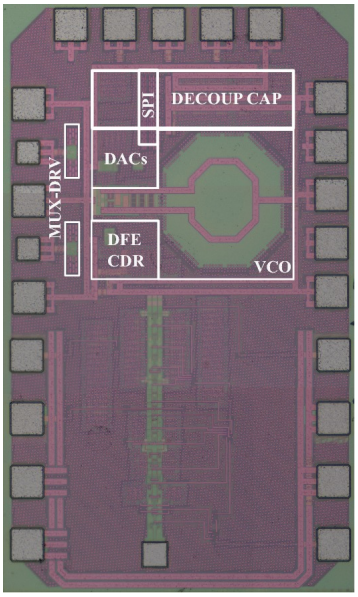


Fig. 6.13: Microphotograph of optical receiver with DFE and CDR chip.

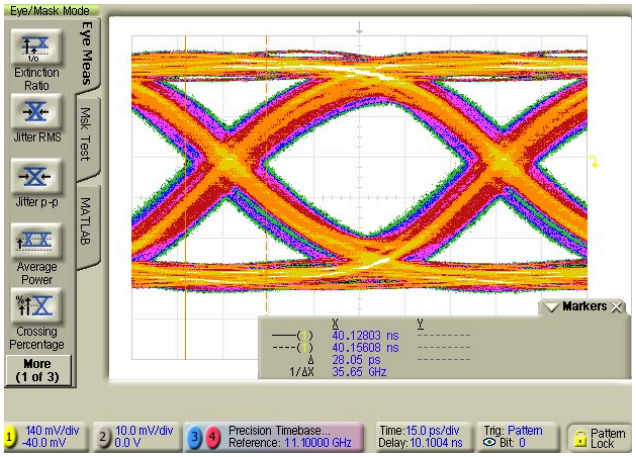


Fig. 6.14: Measured eye at the output of the 2-taps DFE.

7. SUMMARY OF CONTRIBUTIONS

Improving power efficiencies in I/O links is essential to cope with the fast increasing demand for interconnection bandwidth in mobile and server applications. An increasing number of new applications are employing high-speed data links which demands for a lower power consumption and reduced cost. In high data rate I/Os, serializer, pre-emphasis, driver, DFE, CDR and clock distribution circuits take a significant portion of the total power budget because of their complexity and stringent speed requirements. This work presented new circuit techniques that improve performance around energy, speed and area requirements. The following are the summarized contributions of this work:

Introduced implication logic as a design strategy in CMOS. A framework to use implication cells in CMOS as long with the IMP and NIMP cells were designed and measured [54].

Demonstrated a new prescaler architecture using IMP and NIMP cells with further power-efficiency improvement. In a 130nm CMOS technology, the measured results indicated a maximum 17GHz frequency operation. In a 45nm SOI CMOS technology, simulation results show the possibility to build the more compact and more efficient logic-based frequency prescaler in frequencies above the centimeter-wave band [18,55].

Designed and integrated first Sub-1mW prescaler at Ka-band using new logic paradigm jointly with doctor Wu-Hsin Chen. The first implementation of IMP and NIMP logic cells applied to a functional block. The structural complexity of TSPC based prescalers is significantly reduced using non-conventional implication logic cells, achieving one of the best performances among state-of-the-art prescaler designs [19].

Reported the more compact single phase-clock-based divider-by-2/3 prescaler. The proposed circuit technique demonstrated progress toward the application of low-

power and single-phase-clock logic style in mm-wave transceivers using scaled technologies [56].

Formulated the most energy-efficient 4:1 MUX reported operating above 10Gb/s output data rate. The proposed architecture determines feasibility toward the application of low-power and single-phase-clock high-speed I/O links [57].

Designed and integrated the most power-efficient divider-by-2/3 prescaler reported in the literature. Results demonstrated a sub-1mW compact prescaler working in frequencies above 15GHz with 3X lower power consumption compared to prior art in same technology node [58, 59].

Designed and integrated a 30Gb/s transmitter with 4-Tap pre-emphasis for a lossy channel. An energy efficiency of sub-1pJ/b was achieved demonstrating the feasibility to extensively use single-phase clock elements to reduce overall power consumption of a high data rate I/O link [60].

Formulated a new operation scheme of a single-to-differential and duty cycle correction circuits qualified to work at operations above 10GHz. Designed and integrated a proposed divider-by-2 with quadrature output operating at 15GHz [60].

Presented and demonstrated a new D-flip-flop with single-phase clocking operating at 10GHz. The new clock element has a 45% energy reduction compared to the commonly used transmission gate flip-flop in average activity operation. Proposed testing structures to characterize clock elements. The new structures have a reduced area and can be implemented to characterize timing differences in deep-submicron circuits [61].

Demonstrated a new decision feedback equalizer and clock-data recovery circuit. The reuse of elements working at the high-speed paths allow a reduction of power consumption of 30% when compared to separated path elements. The proposed architecture demonstrated the reduction of two times the loading at the input of the DFE, relaxing the requirements of the analog front end circuitry [62].

LIST OF REFERENCES

LIST OF REFERENCES

[1] P. Mayr and et al., "A 90GHz 65nm CMOS Injection-locked Frequency Divider," in *IEEE ISSCC Dig. Tech. Papers*, 2007, pp. 198–199.

[2] W.-H. Chen and B. Jung, "High-Speed Low-Power True-Single-Phase Clock Dual-Modulus Prescalers," *IEEE Trans. Circuits Syst. II*, vol. 58, no. 3, pp. 144–148, Mar. 2011.

[3] Z. Deng and A. Niknejad, "The Speed-Power Trade-Off in the Design of CMOS True-Single-Phase Clock dividers," *IEEE J. Solid-State Circuits*, vol. 45, no. 11, pp. 2457–2465, Nov. 2010.

[4] *The International Technology Roadmap for Semiconductors*, S.I.A Std., 2011. [Online]. Available: http://www.itrs.org

[5] C. Shannon, "A Symbolic Analysis of Relay and Switching Circuits," Master's thesis, MIT, 1937.

[6] A. Whitehead and B. Russell, *Principia Mathematica*. Cambridge at the University Press, 1910.

[7] J. Borghetti, G. Snider, P. Kuekes, J. Yang, D. Stewart, and S. Williams, "'Memristive' switches enable 'stateful' logic operations via material implication," *Nature*, vol. 464, pp. 873–876, April 2010.

[8] E. L. Post, "Introduction to a General Theory of Elementary Propositions," *American Journal of Mathematics*, vol. 43, pp. 163–185, 1921.

[9] P. Tarau and B. Luderman, "Exact Combinational Synthesis and Non-Standard Circuit Design," in *Proceedings of the ACM International Conference on Computing Frontiers*, 2008.

[10] E. Mendelson, *Introduction to Mathematical Logic*. CRC Press, 2010.

[11] A. Tarski, *Introduction to Logic and The Methodology of Deductive Sciences*. Oxford University Press, 1961.

[12] Q. Huang and R. Rogenmoser, "Speed Optimization of Edge Triggered CMOS Circuits for GigaHertz Single-Phase Clocks," *IEEE J. Solid-State Circuits*, vol. 31, no. 3, pp. 456–465, March 1996.

[13] J. Navarro and W. V. Noije, "A 1.6GHz Dual Modulus Prescaler using the Extended True-Single-Phase-Clock CMOS Circuits Techniques," *IEEE J. Solid-State Circuits*, vol. 34, no. 1, pp. 97–102, Jan. 1999.

[14] E. S. Y. Mo, R. Evans, and I. Mareels, "Analysis and design of a 50-GHz 2: CMOS CML static frequency divider based on LC-tank," in *Proc. Eur. Microw. Integr. Circuit Conf.*, 2008, pp. 64–67.

[15] L. Li, P. Reynart, and M. Steyaert, "A 60GHz 15.7mW Static Frequency Divider in 90nm CMOS," in *IEEE ESSCIRC Dig. Tech. Papers*, 2010, pp. 246–249.

[16] K. Scheir and et al., "A 57-to-66GHz quadrature PLL in 45nm digital CMOS," in *IEEE ISSCC Dig. Tech. Papers*, 2009, pp. 494–495.

[17] M. Khrisna, M. A. Do, K. S. Teo, C. C. Boon, and W. M. Lim, "Design and Analysis of Ultra Low Power True Single Phase Clock CMOS 2/3 Prescaler," *IEEE Trans. Circuits Syst. I*, vol. 57, no. 1, pp. 72–82, Jan. 2010.

[18] E. Roa and B. Jung, "A 50GHz 130 μW Inductorless Prescaler in 45nm SOI CMOS Using ETSPC Logic," in *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*, May 2012, pp. 1071 – 1074.

[19] W.-H. Chen, E. Roa, and B. Jung, "A 14.1-GHz Dual-Modulus Prescaler in 130nm CMOS Technology Using Sequential Implication Logic CellsMaterial Implication in CMOS: A New Kind of Logic," in *IEEE RFIC Proceedings*, June 2012, pp. 341–344.

[20] A. Ghilioni, U. Decanis, E. Monaco, A. Mazzanti, and F. Svelto, "A 6.5mW Inductorless CMOS Frequency Divider-by-4 Operating up to 70GHz," in *IEEE ISSCC Dig. Tech. Papers*, 2011, pp. 282–283.

[21] G. Ono, K. Watanabe, T. Muto, K. Fukuda, N. M. nd Tatsuya Saito, and S. Nishimura, "A 10:4 MUX and 4:10 DEMUX Gearbox LSI for 100-Gigabit Ethernet Link," *IEEE J. Solid-State Circuits*, vol. 46, no. 12, pp. 3101–3112, Dec. 2011.

[22] T. Chalvatzis, K. Yau, R. Aroca, P. Schvan, M.-T. Yang, and S. Voinigescu, "Low-Voltage Topologies for 40-Gb/s Circuits in Nanoscale CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 07, pp. 1564–1573, Jul. 2007.

[23] K. Fukuda, H. Yamashita, G. Ono, E. Suzuki, T. Takemoto, F. Yuki, and T. Saito, "A 12.3mW 12.5Gb/s Complete Transceiver in 65nm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, 2010, pp. 368–369.

[24] T. Dickson, E. Laskin, I. Khalid, and S. Voinigescu, "A 72 Gb/s $2^{31} - 1$ PRBS generator in SiGe BiCMOS Technology," in *IEEE ISSCC Dig. Tech. Papers*, 2005, pp. 342–345.

[25] E. Laskin and S. Voinigescu, "A 60 mW per Lane, 4 x 23-Gb/s $2^7 - 1$ PRBS Generator," *IEEE J. Solid-State Circuits*, vol. 41, no. 10, pp. 2198–2208, Oct. 2006.

[26] M.-S. Chen and C.-K. K. Yang, "A Low-Power Highly Multiplexed Parallel PRBS Generator," in *Proceedings of the Custom Integrated Circuits Conference (CICC)*, 2012, pp. 326–328.

[27] A. Yazdi and M. Green, "A 40Gb/s Full-Rate 2:1 MUX in 0.18um CMOS," in *IEEE ISSCC Dig. Tech. Papers*, 2009, pp. 362–363.

[28] D. Kehrer and H.-D. Wohlmuth, "A 30Gb/s 70mW One-Stage 4:1 Multiplexer in 0.13 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 39, no. 07, pp. 1140–1147, Jul. 2004.

[29] W.-Y. Tsai, C.-T. Chiu, J.-M. Wu, S. Hsu, Y.-S. Hsu, and Y.-F. Tsao, "A Novel Low Gate-Count Serializer Topology with Multiplexer-Flip-Flops," in *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*, May 2012, pp. 245–248.

[30] C. Menolfi, J. Hertle, T. Toifl, T. Morf, and M. Kossel, "A 28Gb/s Source-Series Terminated TX in 32nm CMOS SOI," in *ISSCC Dig. Tech.*, 2012, pp. 334–335.

[31] R. Westcott, "Testing Digital Data Transmission in Systems," U.K. Patent 1281390, Tech. Rep., 1972.

[32] M.-S. Chen, Y.-N. Shih, C.-L. Lin, and J. Lee, "A Fully-Integrated 40-Gbs/s Transceiver in 65-nm CMOS Technology," *IEEE J. Solid-State Circuits*, vol. 47, no. 3, pp. 627–640, Mar. 2012.

[33] Y. Lu, K. Jung, Y. Hidaka, and E. Alon, "Design and Analysis of Energy-Efficient Reconfigurable Pre-Emphasis Voltage-Mode Transmitters," *IEEE J. Solid-State Circuits*, vol. 48, pp. 1898–1909, Aug. 2013.

[34] A. Hafez, M.-S. Chen, and C.-K. K. Yang, "A 32-t-48Gb/s Serializing Transmitter Using Multiphase Sampling in 65nm CMOS," in *ISSCC Dig. Tech.*, 2013, pp. 38–39.

[35] J. Yuan and C. Svensson, "New Single-Clock CMOS Latches and Flipflops with Improved Speed and Power Savings," *IEEE J. Solid-State Circuits*, vol. 32, pp. 62–69, Jan. 1997.

[36] V. Stojanovic and V. Oklobdzija, "Comparative Analysis of Master-Slave Latches and Flip-Flops for High-Performance and Low-Power Systems," *IEEE J. Solid-State Circuits*, vol. 34, pp. 536–548, Apr. 1999.

[37] M. Alioto, E. Consoli, and G. Palumbo, "Flip-Flop Energy/Performance Versus Clock Slope and Impact on the Clcok Network Design," *IEEE Trans. Circuits Syst. I*, vol. 57, no. 6, pp. 1273–1286, Jun. 2010.

[38] N. Weste and D. Harris, *CMOS VLSI Design, A Circuit and Systems Perspective*, 4th ed. Addison-Wesley, 2011.

[39] S. Heo and K. Asanovic, "Power-Optimal Pipelining in Deep Submicron Technology," in *ACM Proceedings of ISLPED*, August 2004.

[40] C. Giacomotto, N. Nedovic, and V. Oklobdzija, "The Effect of the System Specification on the Optimal Selection of Clocked Storage Elements," *IEEE J. Solid-State Circuits*, vol. 42, pp. 536–548, Jun. 2007.

[41] M. Alioto, "Ultra-Low Power VLSI Circuit Design Demystified and Explained: Tutorial," *IEEE Trans. Circuits Syst. I*, vol. 59, no. 1, pp. 3–29, Jan. 2012.

[42] N. Nedovic, W. Walker, and V. Oklobdzija, "A Test Circuit for Measurement of Clocked Storage Elements Characteristics," *IEEE J. Solid-State Circuits*, vol. 39, pp. 1294–1304, Aug. 2004.

[43] C. K. Teh, T. Fujita, H. Hara, and M. Hamada, "A 77% Energy-Saving 22-Transistor Single-Phase-Clocking D-Flip-Flop with Adaptive-Coupling Configuration," in *IEEE ISSCC Dig. Tech. Papers*, 2011, pp. 338–339.

[44] E. Consoli, M. Alioto, G. Palumbo, and J. Rabaey, "Conditional Push-Pull Pulsed Latches with 726fJ.ps Energy-Delay Product in 65nm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, 2012, pp. 482–483.

[45] H. Partovi, A. Yeung, L. Ravezzi, and M. Horowitz, "A 3-stage Pseudo Single-phase Flip-Flop Family," in *Symposium on VLSI Circuits Dig. Tech. Papers*, June 2012, pp. 172–173.

[46] N. Kawai, S. Takayama, J. Masumi, N. Kikuchi, Y. Itoh, and Y. Tanaka, "A 77% Energy-Saving 22-Transistor Single-Phase-Clocking D-Flip-Flop with Adaptive-Coupling Configuration," in *IEEE A-SSCC Dig. Tech. Papers*, 2011, pp. 338–339.

[47] S. Parikh, T. K. ad Yasuo Hidaka, and J. Yamada, "A 32Gb/s Wireline Receiver with a Low-Frequency Equalizer, CTLE and 2-Tap DFE in 28nm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, 2013, pp. 28–29.

[48] A. Agrawal, J. Bulzacchelli, T. Dickson, Y. Liu, and D. Friedman, "A 19-Gb/s Serial Link Receiver with Both 4-Tap FFE and 5-tap DFE Functions in 45-nm SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 47, no. 12, pp. 3320–3331, Dec. 2012.

[49] K. Kaviani, T. Wu, A. Amirkhany, and C. Yuan, "A Tri-Modal 20 Gb/s/link Differential/DDR3/GDDR5 Memory Interface," in *Symposium on VLSI Circuits Dig. Tech. Papers*, June 2011, pp. 234–235.

[50] H. Kimura, P. Aziz, T. Jing, Ashutosh, S. Kotagiri, and F. Zhong, "28Gb/s 560mW Multi-Standard SerDes with Single-Stage Analog Front-End and 14-Tap Decision Feedbac Equalizer in 28nm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, 2014, pp. 38–39.

[51] J. Bulzacchelli, C. Menolfi, T. Beukema, D. Storaska, T. Toifl, and D. Friedman, "A 28-Gb/s 4-Tap FFE/15-Tap DFE Serial Link Transceiver in 32-nm SOI CMOS Technology," *IEEE J. Solid-State Circuits*, vol. 47, pp. 3243–3248, Dec. 2012.

[52] N. Kocaman, S. Fallahi, M. Kargar, and A. Momtaz, "An 8.5-11.5Gb/s SONET Transceiver with Referenceless Frequency Acquisition," in *Proceedings of the Custom Integrated Circuits Conference (CICC)*, 2012.

[53] J. W. Jung and B. Razavi, "A 25-Gb/s CDR in 90-nm CMOS for High-Density Interconnects," in *Symposium on VLSI Circuits Dig. Tech. Papers*, June 2012, pp. 113–114.

[54] E. Roa, W.-H. Chen, and B. Jung, "Material Implication in CMOS: A New Kind of Logic," in *ACM Proceedings of DAC*, June 2012.

[55] ——, "A Family of High Efficient Frequency Prescalers using Non-Standard High-Speed Logic in 0.13um CMOS," *IEEE Trans. VLSI Syst.*, Submitted.

[56] ——, "A Polymorphic 15GHz/mW Inductor-less Frequency Prescaler using Implication Logic in 0.13 μm CMOS," in *Radio Frequency Integration Technology (RFIT), 2012 IEEE*, November 2012, pp. 71 – 73.

[57] E. Roa and B. Jung, "A 40Gb/s 860μW Single-Phase 4:1 Multiplexer in 45nm CMOS," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, May 2013.

[58] E. Roa, W.-S. Chen, W.-F. Loke, and B. Jung, "A 36GHz/mW Single-Phase Prescaler using Implication Logic in 0.13 μm CMOS," in *Radio Frequency Integrated Circuits Symposium (RFIC), 2013 IEEE*, June 2013, pp. 341 –344.

[59] E. Roa, W.-H. Chen, and B. Jung, "Analysis and Design of Single-Phase Frequency Dividers," *IEEE J. Solid-State Circuits*, to be submitted.

[60] E. Roa and B. Jung, "A 30Gb/s sub-1pJ/b source series transmitter with 4-tap pre-emphasis," *IEEE Trans. Circuits Syst. II*, to be submitted.

[61] E. Roa, W.-H. Chen, and B. Jung, "A Truly Single-Phase 17T Flip-Flop with 40% Power Reduction in a 20-FO4 deep pipeline," *IEEE Trans. Circuits Syst. II*, to be submitted.

[62] E. Roa and B. Jung, "A 22Gb/s 7.8mW Merged Decision Feedback Equalizer and Clock Data Recovery circuit in 65nm CMOS," *IEEE Trans. Circuits Syst. I*, to be submitted.

VITA

VITA

Elkim Roa received the B.S. degree in electrical engineering from Universidad Industrial de Santander, Colombia, in 1999, and the M.S.E.E. degree from University of Sao Paulo, Sao Paulo, Brasil, in 2003. In 2004 he joined the School of Electrical Engineering at Universidad Industrial de Santander, Colombia as a lecturer. There he founded an integrated circuits design group, specializing in design methodologies and low-power circuits. In August 2009, he joined the analog group at Purdue University where he worked in RF and high-speed wireline communication circuits. Mr. Roa was the recipient of the Fulbright 2009-2011 scholarship. His research interest include energy-efficient high-speed communication circuits, clock generation circuits, mixed-signal circuits, and design methodologies. He was an engineering intern at Samsung, Dallas, TX from January 2013 to July 2013, working on mixed-signal circuit design and verification. From June 2014 to September 2014 he was and enginneering intern at Rambus Inc., Sunnyvale, CA developing analog front-end circuits for high-speed link interfaces.